DOCUMENT RESUME

ED 457 376 CE 082 416

AUTHOR Thomson, P.; Saunders, J.; Foyster, J.
TITLE Improving the Validity of Competency-Based Assessment.
INSTITUTION National Centre for Vocational Education Research, Leabrook (Australia).
ISBN ISBN-0-87397-714-9
PUB DATE 2001-00-00
NOTE 110p.
AVAILABLE FROM National Centre for Vocational Education Research, 252 Kensington Road, Leabrook, South Australia 5068, Australia (cat no. 621; $33 Australian). Tel: 08 8333 8400; Fax: 08 8331 9211; e-mail: vet_req@ncver.edu.au; Web site: http://www.ncver.edu.au. For full text: http://www.ncver.edu.au/research/proj/nr9040.pdf.
PUB TYPE Information Analyses (070) -- Reports - Research (143)
EDRS PRICE EDRS Price MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Competence; Criterion Referenced Tests; Definitions; Developed Nations; Educational Research; Evaluation; Evaluation Methods; Evaluation Needs; *Evaluation Problems; *Evaluation Research; Foreign Countries; *Performance Based Assessment; Postsecondary Education; Secondary Education; Technical Education; *Test Interpretation; *Test Validity; *Vocational Education
IDENTIFIERS *Australia

ABSTRACT

A study was conducted to review the concept of validity in the context of vocational education and training (VET) in Australia. The study included a review of literature, case studies, presentation of key findings, and recommendation of a tool to guide assessors. An eight-faceted approach to validity, based on Nitko, was developed, with each approach using different kinds of evidence (content, internal coherence, external coherence, reliability, practicality, generalizability, consequential) that may be appropriate within a competency-based approach to assessment. The study showed that the eight types of evidence were, with a single exception (consequential evidence), regarded as important by the study participants. The study finds that the approach to validity indicated by Nitko can be fruitful in improving competency-based assessments. Other findings include the following: (1) some practitioners were reluctant to allow any scrutiny of their assessment practices; (2) records storage and retrieval were sometimes not well-developed; (3) there was no uniform definition of "recognition of prior learning" or "recognition of current competency"; (4) the use of holistic assessment left questions of validity open; and (5) attitude seems to be a factor in employment and on-the-job success but was beyond the scope of the study. A revised definition of validity was proposed that sees validity as "the extent to which the interpretation and use of an assessment outcome can be supported by evidence." (Contains 30 references.) (KC)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 457 376

NCVER

Improving the validity of competency-based assessment

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

S. Hayman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

P Thomson

J Saunders

J Foyster

BEST COPY AVAILABLE

CO82 416

Improving the **validity** of **competency-based assessment**

P Thomson

J Saunders

J Foyster

NCVER

© Australian National Training Authority, 2001

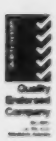
This work has been produced by the National Centre for Vocational Education Research (NCVER) with the assistance of funding provided by the Australian National Training Authority (ANTA). It is published by NCVER under licence from ANTA. Apart from any use permitted under the Copyright Act 1968, no part of this publication may be reported by any process without the written permission of NCVER Ltd. Requests should be made in writing to NCVER Ltd.

The views and opinions expressed in this document are those of the author/project team and do not necessarily reflect the views of the Australian National Training Authority.

ISBN 0 87397 714 9 print edition
0 87397 715 7 web edition
TD/TNC 66.14

Published by NCVER Ltd
ABN 87 007 967 311

252 Kensington Road, Leabrook SA 5068
PO Box 115, Kensington Park SA 5068, Australia
www.ncver.edu.au



Contents

Executive summary 5

Introduction 9

An overview of the literature on validity 9
Construct validity 11
Methodology 12
Focus of study 12
Case studies 12

Validity evidence: Perspectives, opinions and importance 14

Different perspectives on validity evidence 14
Review of assessments of case study participants 16

Assessment review summary 17

Matching industry standards components against assessment 17
Requirements of industry assessment guidelines 19
Assessment instruments: Components 21
Opinions of the case study participants on types of validity evidence 22
Interview summary 22
Relative importance of the eight types of validity evidence 31

Key findings 33

Reluctance to participate 33
Process of auditing 33
Use of 'RPL' 34
Relevance of facets of validity to CBA 36
Workplace vs college assessment 36
Graded assessment vs CBA 37
Attitudinal factors 38
Practical help: A diagnostic tool for improving the validity of assessment 39

References 40

Appendix A 41

Appendix B 52

Appendix C 92

Appendix D 103

Executive summary

This study considers the status of validity in the context of vocational education and training (VET) in Australia. This has involved reviewing the literature, reporting on case studies, presenting key findings and recommending a tool to guide assessors.

The study reports that while validity is an issue that has been considered in previous Australian studies, the approach used has been based upon early models developed in the United States that are now believed to be inadequate. The study therefore turns to more recent approaches to validity to examine their suitability for the Australian environment of competency-based assessment.

One of these approaches takes validity as a unitary entity which, nevertheless can be viewed in a variety of ways: it presents various aspects or facets to which users of assessment outcomes might appeal in seeking to establish the soundness (or otherwise) of the interpretations they make. Thus the focus is shifted from the *validity of an individual assessment instrument* to the broader issue of *the validity of the interpretation and use of an assessment outcome*. The eight facets of validity proposed by Nitko (1996) are the focus of the study.

This approach to validity is examined in the context of the following questions:

- Are these proposed facets of validity meaningful for the competency-based assessment approach now used in Australia's vocational education and training system?
- For a sample of practitioners, to what extent are these expanded notions of validity already familiar?
- If the notions are familiar, are there clear benefits for assessments to be derived from an acceptance and use of this approach to validity?
- Can a 'diagnostic' tool be devised to facilitate use of this approach to validity?

Two groups were identified to be included in the study: organisations that both assess and train assessors at certificate and diploma level, and companies within the retail industry. Six case studies were carried out.

Each case study involved:

- matching the assessment procedures used against industry standards
- study of the organisations' assessment guidelines and instruments and judging them against published national guidelines and assessment component requirements
- interviews with representatives of participating organisations
- preparation of confidential evaluations of assessment procedures (returned to participants and not included in the report)

- preparation of final summaries of each organisation's input (approved by the participating organisations)

The eight-facet approach to validity leads to the identification of eight different kinds of evidence. The report then shows how each of these kinds of evidence, namely:

- representativeness of *content*
- relationships among the assessment tasks *(internal coherence)*
- relationships of assessment results to other variables *(external coherence)*
- *reliability* of assessors and of the assessment over time (stability)
- coverage of thinking skills and processes (*substantive* evidence)
- cost, efficiency and *practicality* features
- *generalisability* for different types of applicant and under different conditions
- value of intended and unintended consequences (consequential evidence)

may in principle be appropriate within a competency-based approach to assessment.

The next step was to develop an interview schedule to explore perceptions of assessment as they presently existed within the two groups being studied. The interview schedule was designed to be carried out in stages, and in association with a consideration of the actual assessment procedures and instruments used in each organisation. The interview schedule focussed on gathering information related to the current use of evidence of the eight types identified by Nitko.

For example, in dealing with the first kind of evidence, the following questions were asked. Do the assessments used in your company/organisation cover all or only some of the content of your training program? Can you say with confidence that your assessments consistently reflect work practice? Do you think any additional assessment is necessary? (If so, what would this be?) Could any assessments or parts of assessment be omitted? Does the emphasis or balance in the assessment match the emphasis on the job? Is the assessment 'up to date'? Are the assessment tasks worthwhile in themselves? (For example, do they contribute to learning?)

That these questions are *not* Dorothy Dixers (which would make the exercise useless) is demonstrated by the fact that only the last of these six questions led to the same answer in all six case studies.

A table sets out the details of each case study in summary, organised in terms of the eight types of validity evidence.

The study showed that the eight types of evidence suggested by Nitko were, with a single exception, regarded as important by the participants. The exception, *consequential evidence*, relates, in part, to the impact of assessment on 'third parties' and would therefore be expected to be less apparent to the organisations being studied.

Based on the information gathered in the case studies, a short self-administered questionnaire with supporting advice has been prepared for assessors, and appears

as appendix A in the report. This diagnostic tool includes practical illustrations that emphasise the importance of the various kinds of validity evidence.

The study presents firm evidence that the approach to validity indicated by Nitko can be fruitful in improving competency-based assessment. The report shows that competency-based assessment is *not* self-validating, and that this is already recognised by industry.

Several other findings are reported:

- There was evidence that some practitioners were reluctant to allow any scrutiny of their assessment practices. Although this was limited in extent, it was nevertheless a cause for concern.
- Because this study required consideration of assessment records kept by participants, it was notable that in some cases, the storage of records and access to them were not well-developed, to the extent that the capacity of the records to be audited was in doubt.
- Participants used ideas of 'recognition of prior learning' and 'recognition of current competencies' in various ways, indicating that there was no uniform understanding of this area.
- The use of 'integrated competency assessment' (holistic assessment), while valuable, appeared to raise some issues regarding validity that have yet to be resolved.
- While the issue of grading arose in the course of the study, its role is complex and issues of validity would vary depending upon use.
- One influence on employment and on-job success that also has an impact on validity could not be considered in the study because it appears to be an unstated factor. This is the matter of attitude. In some cases this may be a major factor, and its absence from consideration in training packages is a cause for concern. This was beyond the scope of the present research, but it did place some limitations upon the results produced by this study.

The study has been able to demonstrate that, within the industrial areas studied, the proposed expanded approaches to validity are already generally regarded as important. The use of the broader notion of validity would allow a clearer understanding of practical issues. A diagnostic tool has been developed that will facilitate the development of such understanding.

Finally, as a consequence of this broader approach, a revised definition of validity is proposed that sees it as *the extent to which the interpretation and use of an assessment outcome can be supported by evidence.*

Introduction

This report, *Improving the validity of competency-based assessment* provides an overview of the literature on validity, outlines the findings of a series of case studies undertaken by the researchers, identifies the key findings from the study, and proposes a diagnostic tool for use in the training and development of assessors in vocational education and training.

An overview of the literature on validity

There is a relatively small amount of Australian research on validity, especially in the area of competency-based assessment. In their review of the literature on assessment of competency-based training (CBT) in Australia, Toohey et al. (1995) make only indirect references to validity issues. Furthermore, most sources (NTB 1992; Hager et al. 1994; Cropley 1995) define 'validity' along the lines that it is: *The extent to which the assessment method measures what it is supposed to measure.*

This definition is essentially a contraction of a much earlier attempt by Lindquist (1942) to define validity in which he stated:

> *The validity of a test may be defined as the accuracy with which it measures that which it is intended to measure, or as the degree to which it approaches infallibility in measuring what it purports to measure.* (p.213)

However, as Ebel (1965) pointed out, this conception of validity has shortcomings because it raises the question *What should the test measure?* Definitions which raise substantial questions of this nature are most unsatisfactory.

It is therefore disappointing to learn that the recently published training package of the National Assessors and Workplace Trainers' Body (NAWTB 1999) defines validity in a similar manner as that proposed by Linquist, namely:

> *A valid assessment assesses what it claims to assess; evidence collected is relevant to the activity and demonstrates that the performance criteria have been met.* (p.18)

The training package definition shares the shortcomings of Lindquist's earlier version and further compounds this by adding what presumably is an 'example' of how to conduct valid assessment namely, *the collecting of relevant evidence that performance criteria have been met*. This addition basically has the training package defining validity as *content validity;* that is, validity concerned with showing how well the content of the assessment samples the situations (performance criteria) about which conclusions are to be drawn.

Elsewhere in the package under 'rules of evidence', when describing 'valid evidence', the following statement appears:

> *Evidence of competence must cover the broad range of knowledge, skills and the application of such knowledge and skills specified in the Assessment and Workplace Training Competency Standards, assessors need to ensure that the evidence collected focuses on the appropriate knowledge and skills specified in the Performance Criteria and Evidence Guides.* (p.21)

Again, we can see the authors of the package are clearly relating 'validity' with 'content', in particular, the content of competency standards as specified by the performance criteria. But 'content validity' is by no means the only aspect of validity considered in the literature. Typically, a classification of validity components comprises the following:

- *content validity* which is evaluated by showing how well the content of the test samples the class of situations or subject matter about which conclusions are to be drawn
- *criterion-related validity* which is evaluated by comparing the test scores with one or more external variables (called criteria) considered to provide a direct measure of the characteristics or behaviour in question. Criterion-related validity's two sub-divisions are:
  - *predictive validity* which indicates the extent to which an individual's future level on the criterion is predicted from prior test performance
  - *concurrent validity* which indicates the extent to which the test scores estimate an individual's present standing on the criterion
- construct validity which is evaluated by investigating what qualities a test measures, that is, by determining the degree to which certain explanatory concepts or constructs account for performance on the test (adapted from Messick 1989, p.16)

Recent developments in our understanding of validity show that this multi-dimensional approach is nevertheless outdated and that a new approach is required.

The new approach is one which replaces the notion of validity being made up of elements by an approach which gives something like 'construct validity' as described above, the central or unifying role.

> *Almost any kind of information about a test can contribute to an understanding of its construct validity, but the contribution becomes stronger if the degree of fit of the information with the theoretical rationale underlying score interpretation is explicitly evaluated.* (Messick 1989, p.17)

Although not reflecting this new approach, the NAWTB training package does recognise the fundamental importance of validity in assessment. The package includes validity as one of the four technical principles that underpin assessment, namely, validity, reliability, flexibility and fairness. However, as will be seen, it is possible to have a wider conception of validity which subsumes principles of reliability, flexibility and fairness. Validity is the main assessment game—some would say the only game!

Our project was based on this broader approach to validity which sees it as *the extent to which the interpretation and use of an assessment outcome can be supported by evidence.*

This in effect means that validity is not an intrinsic property of assessment instruments, but rather it refers to the soundness of the interpretations and uses of the outcomes of an assessment. In other words, the project adopts a different definition of validity than from commonly used in Australia. Gillis and Bateman (1999) were among the first in Australia to recognise the new approach. They do not however, embrace the unitary nature of validity that follows from the new definition.

More specifically, the project has drawn on the work of Nitko (1996) who proposed eight types of validity evidence that can be addressed for any assessment activity. These are set out below; the project has tested each of these in the context of competency-based assessment.

Nitko's eight types of validity evidence:

1. content representativeness and relevance (called *content evidence*)
2. relationships among the assessment tasks or parts of the assessment (called *internal coherence evidence*)
3. relationships of assessment results to the results of other variables (called *external consistency evidence*)
4. reliability over time, assessors and content domain (called *reliability evidence*)
5. types of thinking skills and processes required (called *substantive evidence*)
6. cost, efficiency, practicality, instruction features (called *practicality evidence*)
7. generalisabiltiy over different types of people, under different conditions (called *generalisability evidence*)
8. value of the intended and/or unintended consequences (called *consequential evidence*)

(Nitko 1996)

A complete review of the literature is provided in appendix C of the report.

Construct validity

The pre-eminent place that the new approach to validity gives to construct validity was the subject of some discussion amongst the participants early in the project. In particular, there were concerns over just what constituted a 'construct' in competency-based assessment.

Until very recently a 'construct' in educational and psychological assessment was understood to be some trait or characteristic that was not directly observable. The construct had to be inferred from an observation. Therefore 'competence' was a construct inferred from observations of performance. This view of constructs was not without its critics, but the American Education Research Association's (AERA) (1999) standards have overcome the problem to some extent by departing from the historical use of the term 'construct'. The standards adopt a broader definition of a construct as being *any concept or characteristic that an assessment is designed to measure*. Under this definition, competence and performance are both constructs, and in a competency-based system such as exists in Australia, all assessment evidence can be related to a concept or characteristic and hence a construct. This approach has been adopted by the project team.

Methodology

The project methodology involved:

- preparation of a literature review
- consultations with industry, enterprises and training providers to identify suitable organisations for case studies
- development of interview protocols (appendix D) and associated survey instruments
- a survey of individuals who had the responsibility for assessment in the participating companies and training provider organisations
- an investigation of existing approaches for interpreting and using validity evidence in the participating companies and training provider organisations

The four main data-gathering activities were:

- the review of the literature relating to current approaches to validity
- the case study investigations of the participating organisations
- the review of a sample of trainer/company assessment instruments (see Assessment Review Process—appendix D)
- the interviews with assessors or company officers (see appendix D—this is the instrument used with assessors in company registered training organisations (RTOs), as well as assessors from the assessor and workplace trainer group). The questions and responses can also be found in the 'interview summary' in the next section

Focus of study

During the planning stage it became clear that the people to target in the study were assessment practitioners. Individuals with assessor qualifications gained through the Training Package for Assessment and Workplace Training (NAWTB 1999) therefore became one focus of the study. This led to the decision to include organisations which trained assessors at certificate and diploma level.

The second focus was to provide the study with an industry base and here the choice was the retail industry. Many medium-to-large size companies in the retail industry are currently committing substantial resources to staff assessment and the issues this presented were highly relevant to the purpose of the study.

Case studies

Six companies and organisations covering the retail sector and using assessors with qualifications from the National Assessors and Workplace Trainers Body took part in the study. These were:

- Woolworths Supermarkets
- McDonald's (Australia) Ltd
- National Pharmacies

- David Foreman and Associates
- Logistics Training Unit, Douglas Mawson Institute of TAFE
- Retail Training, Regency Institute of TAFE

Reviews of the assessment procedures of each organisation were conducted to determine how faithfully the assessment processes reflected the industry standards as laid down in the training packages (or company program/curriculum). The review process involved members of the research team working through the components of the standards and cross-checking these against the assessment 'instruments' being used by the assessors to judge competency.

As well as this matching exercise, the researchers made judgements about the assessment guidelines and the components of the assessment instruments themselves. In the case of the assessment guidelines, a judgement was made against the published guidelines (that is, the guidelines of the retail industry and National Assessors and Workplace Trainers Body). In the case of the assessment instruments components, judgements were made by examining the instruments against the component requirements set down in the *Training package for assessment and workplace training* (p.15).

In addition, representatives of the participating organisations were interviewed to obtain information about the validity of their assessment procedures.

The information obtained from the above activities was brought together in six separate case study reports that are included in appendix B.

Participating companies were also provided with a separate confidential evaluation of their assessment procedures. These have not been included in this report.

In the next section we look at the different perspectives on validity and summarise the opinions of representatives of the organisations taking part in the case studies. This information is then analysed to determine the relative importance currently being given to the eight types of validity evidence in vocational education and training in Australia.

Validity evidence: Perspectives, opinions and importance

Different perspectives on validity evidence

As already noted, validity *is not an intrinsic property of assessment instruments*, but rather it refers to the soundness of the interpretations and uses of the results of an assessment. Furthermore, there has been an evolution in the meaning of 'validity': 'this evolution has involved a shift from validity being broken into distinctive types and sub-types (that is, content validity, predictive validity etc.) to a single or unitary notion of validity.

The need for concern about the soundness of interpretations arises from the complex environment within which assessments are generally made (for example, by one or more assessors), and in particular, the different perspectives from which interpretations may be made. There are at least three distinct perspectives on an assessment activity: that of the person being assessed, that of the person (or persons) carrying out the assessment, and that (or those) of a third party (or third parties) who observe and/or make use of the assessment outcome. For convenience of reporting we will put these 'parties' into a workplace context and identify them as *trainees*, *assessors*, and *employers*. (In some cases, however, this is an oversimplification, and we will need to broaden the notion of 'employer' to embrace 'all of those affected by the working lives of trainees'.)

These different parties may have varying views on the soundness of interpretation of an assessment outcome and we must attempt to determine whose perspectives and views are reasonable and relevant *in the given context*.

Therefore in deciding whether a particular assessment is valid in context, we need to understand what evidence the parties will want to take into account before becoming concerned or wanting to challenge the assessment outcome.

Nitko (1996) suggests that there are eight kinds of evidence that might enter into any formal argument about the soundness of interpretation and use of an assessment result. We can see how concerns about these might arise in the minds of the *trainees*, the *assessors*, and the *employers*. In this study we have been particularly interested in the extent to which those interviewed have addressed the issue of the need for different kinds of evidence as described by Nitko. We can consider these types of evidence in terms of the possible challenges that an evaluator might point to when reviewing the validity of an assessment process.

- The *trainee* might challenge the use of an assessment result if the content of the assessment instrument does not fairly reflect the content of the course studied. An *assessor* would be wise to verify, before using an assessment instrument, that such a challenge could not be reasonably made. An *employer* should be able to examine documents and similar artifacts to establish the

truth or otherwise of any challenge. Nitko describes this kind of evidence as relating to *representativeness of content*.

- A *trainee* might reasonably challenge the use of a particular assessment task on the grounds of its being significantly different from other assessment tasks in the same course (in particular, if it is much more difficult or much easier than other assessment tasks). (This kind of variation would also be of concern to the *assessor* who has anticipated no difference from other assessments.) An *employer* can determine whether or not this is true by simple inspection of assessment results or instruments. Nitko refers to this aspect as a requirement for *internal coherence*.
- An *employer* might challenge the use of an assessment instrument on the grounds that it appears to have insufficient relationship to the purpose for which it was designed; for example, if it is used in the assessment of potential salespeople, then those who succeed on the task *ought* to turn out to be good at selling. *Trainees* who were not 'successful' might reasonably be expected to lodge objections if the 'successful' trainees turned out to be poor at selling. Nitko identified evidence that the assessment task is relevant as evidence of *external consistency*.
- Both *employers* and *trainees* might challenge assessment interpretations made by *assessors* who show inconsistent patterns of assessment results. It is reasonable to expect assessments to show stability from year to year, or within a given year. Nitko refers to this condition as *reliability* of assessment.
- An *assessor* might be challenged by a third party wishing to interpret assessment results (for example, an employer/supervisor) to prove, say, that the assessment actually involved relevant and significant thinking skills as opposed to simple unthinking repetition of tasks performed mechanically as a result of rote learning. The expectation that an assessment will require the demonstration of such essential skills is described by Nitko as a requirement for *substantive* evidence.
- *Assessor*, *trainee*, and *employer* may all, for various reasons, object to particular approaches to assessment if these are expensive, inefficient, or impractical, since such approaches can produce a range of problems for any party to the assessment exercise. Evidence regarding the *practicality* of an assessment task is therefore identified by Nitko as a distinct kind of evidence required to be considered.
- *Trainees* could be expected to challenge assessment results that depend on (performance-irrelevant) background factors such as gender, class, or ethnicity of either *trainee* or *assessor*, or upon temporary or local special conditions (which would not apply after selection). In this case Nitko's requirement is that the assessment result must be *generalisable*.
- Finally, a challenge to an assessment which might arise from an *observer acting independently of the three 'parties'* when the assessment has consequences for the rest of society which, although it might be unintended, is nevertheless perceived as negative. (For example, many *observers* feel that the end-of-secondary-school examinations have negative effects on learning.) This kind of evidence is regarded by Nitko as being *consequential*.

In this study we have sought to document the extent to which assessors have anticipated possible challenges to their assessments or their use and interpretation of assessment results. Validity should be a component of the training and assessment strategy, rather than being regarded as a technical and academic exercise of little relevance to day-to-day instructional life.

As we indicate above, the consequences of failing to address issues of validity can be substantial (in terms of defensibility) should an evaluation or audit reveal the outcomes of the process to have little merit. Accordingly, prudence demands that designers of assessment systems ensure that the systems are valid in at least the eight senses described above. A practical approach to validity such as that used in this study has distinct advantages over validity arguments based upon theoretical considerations only.

Review of assessments of case study participants

A review of the assessments used by each case study participant was carried out to determine how faithfully the assessment processes reflect the industry standards as laid down in the training packages (or company program/curriculum).

The review process involved members of the research team working through the components of the industry standards and cross-checking these against the assessment 'instruments' which were being used by the assessors to judge competency.

As well as this matching exercise, judgements were made against each of the guidelines for assessments specified in the training packages and, finally, an estimate was made on how well the published components of an industry standard (task skills, management skills, transfer skills etc.) were covered.

Assessment review summary

Matching industry standards components against assessment

The assessment review is outlined in appendix D. The organisations involved are identified by the letters P, Q, R, S, T and U. A summary of the outcomes extracted from the case studies follows. The individual case study reports can be found in appendix B.

Questions	Organisation					
	P	Q	R	S	T	U
1. Performance criteria of the relevant elements of competency 1.1 Are all the performance criteria covered?	All performance criteria are covered (including many organisation specific criteria).	All performance criteria are covered and checked off by the trainee in the training program, but only a sample of the criteria is assessed.	Yes	Not all performance criteria are specifically stated in the assessment documents. However they should be covered if appropriate assessment tasks are chosen.	All performance criteria are covered.	All performance criteria are covered.
2. Range of variables (ROV) 2.1 Do the assessments cover the range of situations and equipment listed?	Yes	An extensive range of variables is covered by the training documentation, however, only a sample is covered in an assessment.	Yes	Yes, provided assessor exercises appropriate discretion in choosing from the suggested assessment tasks.	Assessment tasks provide scope for full ROV to be assessed but actual cover relies on expertise of assessor	Yes

Matching industry standards components against assessment cont.

Questions	Organisation					
	P	Q	R	S	T	U
3. Evidence guide						
3.1 Are all the critical aspects covered?	Yes	Can be — dependent on experience and expertise of the assessor.	Yes	Not fully specified in assessment documents. Assessor would need to refer to the training standards for complete information.	Yes, through repeated observation of trainee.	Yes
3.2 Are underpinning knowledge, skills and attitudes formally assessed?	Yes	Can be — dependent on experience and expertise of the assessor.	Yes	Not fully specified in assessment documents. Assessor would need to refer to the training standards for complete information.	Yes, these aspects are covered in detail using special training materials.	Yes
3.3 Are 'What if ...' questions used?	Yes	Yes	Yes	Yes	Although not specified in the training documents, assessors are expected to verify understanding and transferability.	Yes
4. Key competencies						
4.1 Are key competencies assessed separately or is acquisition of key competencies assumed through 'application' type evidence?	Some key competencies specifically assessed.	Key competencies are not assessed separately, although some aspects of key competencies are specifically tested.	Key competencies are not assessed separately.	Key competencies are not assessed separately.	Key competencies are not assessed separately.	Key competencies are not assessed separately.

Requirements of industry assessment guidelines

Questions	Organisation					
	P	Q	R	S	T	U
1. Has a database of qualifications awarded and statements of attainment issued been developed? If so, how long has it been in use?	National records have been kept for some years.	Assessment records are kept by the organisation on state-based central databases.	Records going back 6 years are available	Yes, a State-based database has been in use for many years.	Assessment records are kept by the organisation on its State-based central database.	Yes, State-based database has been in use for many years.
2. Have the assessment procedures of the company/ organisation ever been audited? If so, when and by whom was this last done?	Yes, as part of 'Quality endorsement' requirements.	Yes, regular intra-organisation reviews and audits are conducted. Auditing by external bodies also takes place.	Yes, in accordance with registration requirements for RTO status most recently in February 2000.	There is no formal auditing of assessment procedures.	Apart from that required for RTO status, no formal auditing takes place.	Although there are no explicit formal audits, there are regular meetings of personnel from this and sister organisations during the year, in which assessment procedures are reviewed.
3. Do the assessments have an appeals process in place as set down in the guidelines?	Yes	Yes	Yes, procedures are documented in the organisation's manual of operation.	Yes, the organisation has an established and documented appeals process.	Appeals may be heard under the organisation's established grievance policy.	Yes, the organisation has an established and documented appeals process.
4. Are recognition of current competencies/ recognition of prior learning (RCC/RPL) procedures allowed as part of competency assessment?	Yes, but the specialised nature of work tasks performed by the organisation's employees makes it difficult to establish equivalence.	Yes, recognition is awarded through successful completion of the relevant assessment.	Yes, RCC and RPL often constitute a significant component of the assessments conducted.	Yes, the organisation distinguishes between RCC and RPL and applies either or both processes depending on the context in which recognition is sought.	Yes	Yes

Requirements of industry assessment guidelines cont.

Questions	Organisation					
	P	Q	R	S	T	U
5. Are 'integrated competency assessments' (ICAs) used? (Refer Retail Training Package)	Yes, at certificate III and above.	No	Yes	ICAs are not used specifically. However, because of their highly holistic nature, the assessments match many of the characteristics of ICAs.	Not at this stage. This aspect is under review.	Yes
6. If ICAs are used do they adopt the guideline principle of multiple assessments, i.e. a minimum of 3 pieces of evidence?	N/A	N/A	Holistic assessment used when appropriate with 3 or more pieces of evidence the norm.	N/A	N/A	Yes
7. Do the assessments gather enough evidence, i.e. are enough contexts covered and is consistency of performance taken into account?	Yes	Yes	Yes	Yes, much of the assessment is on the job, so there is ample opportunity to assess in various contexts and over extended periods of time.	Yes	Yes
8. Are steps taken to authenticate any work samples or other evidence presented for assessment that was not done in the presence of the assessor?	Yes, organisation policy requires assessors to verify and authenticate trainees' work.	Limited authentication; assessors are expected to authenticate unsupervised work presented for assessment. Other unsupervised work is not authenticated.	Yes, this is a particularly important component of assessment because of the large amount of RCC/RPL often used.	Authentication does take place at the discretion of the assessor. However there are no documented requirements or procedures.	There is no formal policy on authentication. This issue is to be looked at in the future.	Yes, more so at the higher levels.
9. Do the assessments have provisions for dealing with the requirements of 'equal opportunity' legislation?	Yes	Yes, the organisation complies with all federal and state legislation relating to equal opportunity.	Yes	Nature of the assessment tasks helps ensure equal opportunity. The organisation has a well established EO policy.	Yes, in accordance with the EO component of the organisation's HR policy.	Yes, the organisation has stringent procedures to ensure equal opportunity requirements are met.

Assessment instruments: Components

Questions	Organisation					
	P	Q	R	S	T	U
1. Do the assessment instruments, taken together, cover task skills (performing at an acceptable level)	Yes	Yes	Yes	Dependent on the experience and expertise of the assessor due to the holistic nature of the assessments.	Yes	Yes
2. Do the assessment instruments, taken together, cover task management skills (managing a number of different tasks that make up a job)	Yes	Yes	Good coverage is possible, but is dependent on the particular project work on which the assessment is based.	Dependent on the experience and expertise of the assessor due to the holistic nature of the assessments.	Yes	Yes
3. Do the assessment instruments, taken together, cover contingency management skills (dealing with the unexpected)	To a limited extent.	Yes	Good coverage is possible, but is dependent on the particular project work on which the assessment is based.	Dependent on the experience and expertise of the assessor due to the holistic nature of the assessments.	Yes	Yes
4. Do the assessment instruments, taken together, cover job/role environment skills (working with others, working as part of a team)	Yes	Yes	Good coverage is possible, but is dependent on the particular project work on which the assessment is based.	Dependent on the experience and expertise of the assessor due to the holistic nature of the assessments.	Yes	Yes
5. Do the assessment instruments, taken together, cover transfer skills (using skills learnt in one situation in a new or different context)	Transfer skills not explicitly covered in assessment documentation.	Transfer skills covered only to a minor extent.	Good coverage is possible, but is dependent on the particular project work on which the assessment is based.	Dependent on the experience and expertise of the assessor due to the holistic nature of the assessments.	Not explicitly covered in assessment documentation. However opportunities for demonstration of skills transfer do exist.	Not specifically covered in assessment documentation, but assessment tasks support transfer of skills.

Opinions of the case study participants on types of validity evidence

Representatives of the organisations in the case studies were interviewed on a range of issues related to the validity of assessment. The aim was to gather information on how they were dealing with the eight types of evidence identified by Nitko (1996), namely:

- representativeness of *content*
- relationships among the assessment tasks *(internal coherence)*
- relationships of assessment results to other variables *(external consistency)*
- *reliability* of assessors and of the assessment over time (stability)
- coverage of thinking skills and processes *(substantive evidence)*
- cost, efficiency and *practicality* features
- *generalisability* for different types of applicant and under different conditions
- value of intended and unintended consequences *(consequential evidence)*

Interview summary

The interview schedule is provided as appendix D and a summary of the six sets of responses follows. The organisations involved are identified by the letters P, Q, R, S, T and U. The individual case study reports can be found in appendix B.

Questions	**Organisation**					
Content	P	Q	R	S	T	U
1. Do the assessments used in your company/ organisation cover all or only some of the content of your training program?	All (with focus on achieving consistency).	All	Follow training package requirements.	All	Only Certificates I to III. Certificates IV and Diploma are covered by outside providers.	Assessments cover all elements and performance criteria of the training programs for certificates I to IV.
2.1 Can you say with confidence that your assessments consistently reflect work practice?	Some assessors could do a better job (see Q15).	Confident	Confidence is obtained through a range of checks of workplace performance and interviews with peers.	Assessments are thought to reflect work practice.	Yes	Yes, however, employers occasionally say some training content is not relevant to their organisation, but the organisation regards this as necessary in order to address everything in the qualification.

Interview summary cont.

Questions	Organisation					
Content cont.	P	Q	R	S	T	U
2.2 Do you think any additional assessment is needed? If so, what would this be?	Could be more theory in workplace assessments.	Could be more attention to keep currency of a skill (i.e. retesting).	No	No additional assessment is needed because everything required is laid down in the competency standard which the assessment matches.	No, the only exception might be introduction of integrated assessment if a suitable compromise can be found.	No
2.3 Could any assessments or parts of assessments be omitted?	No, but there have been attempts to 'streamline'.	Integrated competency assessment is being considered.	Believes all current processes are needed.	No	No	Possibly, the combination of both individual and integrated assessment can result in 'over-assessing' and therefore reduce cost-effectiveness. This is more of a problem at higher levels.
2.4 Does the emphasis or balance in the assessment match the emphasis 'on the job'?	Yes	Balance is problem; some elements need more time than others.	Yes	Yes	Yes	Yes, all assessment is based on what is expected in the job.
2.5 Is the assessment 'up to date'?	Yes	Training & assessment for GST not up to date.	Tailored to customer needs.	Assessments are thought to be up to date.	Yes	Yes, industry-based lecturers and the quality assurance group keep the assessments up to date.
2.6 Are the assessment tasks worthwhile in themselves? (e.g. do they contribute to learning?)	Tasks integrated with learning.	Very worthwhile	Integrated with training.	Yes	Yes	Yes, assessment requires the trainees to think about how to apply the theory they have learnt.

Interview summary cont.

Questions	Organisation					
Internal	P	Q	R	S	T	U
3.1 Where different assessment tasks have some overlap in content, are the assessment outcomes from the different tasks consistent?	Because approach is company specific there is little overlap between assessment tasks.	Outcomes are consistent.	Consistent	Assessments don't have much overlap.	There may be some overlap, such as in communication. No significant differences in results for communication assessments have been noticed.	Generally, if the trainee has been shown to have achieved a competence, that competence is not reassessed.
3.2 Is the interpretation of the outcomes of the assessment tasks unambiguous? What do you do if the outcome from a particular task is ambiguous? (For example, if repeated attempts are allowed, how many repeats do you allow?)	Outcomes unambiguous	'As many repeats as needed' are allowed.	Repeats allowed after a period of further training.	No limit specified.	No limit specified.	Three attempts initially allowed, after which further training is required before subsequent attempts.
External	P	Q	R	S	T	U
4. Are the outcomes of the assessment tasks consistent with other evidence such as work assessments or third-party verification (testimonials)?	Concentration is on assessment task and repeated consistent performance of that task. More than one assessor is involved in each assessment.	Assessors' job is to look for other evidence.	Emphasis is kept on gathering evidence from a variety of sources. When inconsistencies emerge they are investigated and resolved.	All the assessment tasks are in the workbooks where there is provision for supervisors to comment on progress, achievements and any problems. Third-party verification also applies when doing RCC.	Cannot say However, third-party verification (by the supervisor) is intrinsic to the assessment.	In the case of trainees, the organisation has never had an occasion in which the trainee has been judged competent but the employer has refused to 'sign them off'. Generally, supervisor feedback corresponds well with assessment findings.

Interview summary cont.

Questions	Organisation					
External cont.	P	Q	R	S	T	U
5. Do the outcomes of the assessment tasks satisfactorily predict job success, for example through promotion, job/work culture changes, employer satisfaction?	It is more the performance appraisals that do this (these are separate from the workplace assessments).	Not really, unless they make an extra effort the trainees stay in the job they are qualified to do.	Does not follow up later performance. Believes this is employer responsibility.	Yes, employers write letters expressing satisfaction.	Too early to be able to answer (assessments have not been in use long enough).	Yes, employment outcomes are good, many trainees find jobs.
6. Do the outcomes of the assessment tasks satisfactorily predict employment success in terms of further learning (either on or off the job)?	Depends a lot on the trainee. Trainees may have no intention of progressing in the system. Again it is the performance appraisals that count.	See above sometimes, yes, sometimes, no.	Does not get involved with this; it is more an issue for the employer.	Most of the trainees stay on with the employer with which they have been trained. Employers are using success at one level to encourage trainees to move on to the next level.	Too early to be able to answer (assessments have not been in use long enough).	Trainees sometimes achieve competency in level 2 but run into problems at level 3. This is because certificate 2 is very much a hands-on skill-based qualification, whereas certificate 3 jumps into co-ordinating work teams, leadership, writing procedures and the like.
7. Are assessments graded in any way e.g. to identify those who are 'very competent' rather than 'just competent'? How is this done.	Grading is via performance appraisals: • outstanding • excellent • good • need to improve	No grading	No grading, but can provide 'performance appraisal' if required/ requested.	No grades	No grades ...	Yes. But only at certificate IV and above. Grading is not based on marks, but instead on the assessor's judgement of additional criteria.

BEST COPY AVAILABLE

Interview summary cont.

Questions	Organisation					
Reliability	P	Q	R	S	T	U
8.1 Are the outcomes of your assessment independent of the timing of the assessment. Do you have to pick the right time?	Assessment can be announced or unannounced but still important to pick the right time.	Timing is important to catch the skills in workplace e.g. baking at midnight.	Important to find a suitable time, avoid 'dead areas' in shifts.	A person assessed at the beginning of an afternoon shift will perform differently to when tested at the end of a shift.	This should not be an issue. The assessment times are chosen by the student, so the issue largely rests with them.	This can be a problem. For example, if the programs are not completed within the set time the trainees will need to re-enrol at extra cost, so there is some pressure to complete assessments by the end of semester.
8.2 Have you every been forced to conduct an assessment at a time you considered inappropriate?	Trainees are expected to perform at all times.	Timing moved if it doesn't cost (they do not want to use simulations if real thing is available).	Weather can be a factor.	Assessment times are bound, to some extent, by employer requirements.	Not applicable (see above)	Undoubtedly, assessors sometimes conduct assessments when circumstances are not at their best, such as feeling unwell.
9. Are the outcomes of the assessment tasks independent of the assessor(s)? (For example, are you confident that another assessor would come up with the same assessment results as you?)	The specific nature of the assessments ensures the assessors can all assess to the (same or similar) standard.	Confident if it is workplace assessment.	Part of assessor training involves 'moderation' sessions with other assessors to ensure assessment is to agreed standard.	Yes, even if you swap your assessors over, you still get consistent results for individual trainees.	Too early to say. But would expect outcomes to be independent of assessors.	This is always an issue. However in level 2, for example, the co-ordinator and lecturers have all been involved in developing and implementing the assessments (looking at standards and content) and they compare assessment procedures and results. So in effect there is a moderation process operating.

Interview summary cont.

Questions	Organisation					
Reliability cont.	P	Q	R	S	T	U
10. Would it be useful to have equivalent forms of the assessment tasks available? (For example, to have an equivalent form or a knowledge test that you could use?)	No	Probably, yes	Yes, and has developed equivalent forms for some knowledge tests. (This, in part, driven by concerns about security of tests.)	CBA is transparent, students always know what they are expected to do, so the question of equivalent forms of assessment is irrelevant.	Equivalent forms of assessment not seen to be relevant as the assessments are so strongly performance-based. The issue of question banks for assessment of knowledge was considered but not adopted.	Not really. If a trainee does badly, they are given feedback and extra training and are then re-assessed on a similar task. Where a person has not done well, they may be given a new assignment to do as part of their assessment.
Substantive	P	Q	R	S	T	U
11. Do the assessment tasks include the application of thinking skills other than rote learning?	This is an outcome of workplace assessment; customer service on the 'floor' is an example.	Yes (workplace again)	Rote learning does predominate at certificate level, especially level 2, but thinking skills assessed at higher levels.	Yes, dependent on level; because levels 2 and 3 are basically the operative level where you've got skills and a little bit of underpinning knowledge. Once you get to level 4 you've got a lot of under-pinning knowledge to the skills.	Most assessment tasks require the trainee to discuss or report on what they have done. The assessments have tried to get away from merely ticking a box; trainees now need to demonstrate they have the knowledge and can apply it. (There is, however, still a lot of rote learning in some tasks, such as operating a cash register.)	Yes, evidence is usually gathered from three different areas, it is not just a matter of writing down an answer, trainees are observed putting the learning into practice.

Interview summary cont.

Questions	Organisation					
Practicality	P	Q	R	S	T	U
12. Are your company's/ institute's assessments recognised by all relevant groups and individuals (management, trainers, assessors and trainees) as being appropriate?	All company certificates and diplomas are nationally recognised. Holders of company's qualifications are sought after by other companies.	Qualifications recognised nationally.	Assesses to national requirements.	Yes, by the industry in general and by the relevant industry associations.	No, there is a reluctance of some other sections of the industry using different training and assessment programs to accept the company's outcomes.	Most assessments are based on the WRAP's* guidelines to assessment. Most lecturers (conducting assessments) have recent retail industry experience or are currently in the industry and so assessments are recognised by industry. Letters have been received from employers confirming what is being covered is relevant.
13. Have you conducted any investigations to check how cost-effective your assessments are?	No, but performance is constantly under scrutiny.	Informal only, concentration is on training rather than assessment costs.	Not applicable as operating as fee for service.	Have never conducted any formal investigations.	No	No, however, it is felt that the combination of individual and integrated competency assessment in the retail training package is really 'over-assessing' and therefore not very cost-effective.

Note: * Wholesale Retail and Personal Services

Interview summary cont.

Questions	Organisation					
Practicality cont.	P	Q	R	S	T	U
14. What percentage of your assessment tasks require you to assess on a one-to-one basis?	100%	Approximately 100%	Approximately 90% Group assessment only used with knowledge tests.	About 90% of it. The only area not assessed on a one-to-one basis is theory which is done in a classroom —limited to 10 trainees at a time.	As far as supervisors and managers are concerned, the majority of assessments are done on a 1:1 basis. The supervisor usually has to observe, review or question the trainees individually.	It is estimated that about 90% of assessment is competency-based.
15. Are you satisfied with the assessment skills of the available assessors, in terms of their ability to implement the assessment procedures?	On the whole yes, but room for improvement.	Currency of skills is problem. Those assessing regularly are OK, but occasional assessors not so good.	Believes the skills of group are first-class.	Yes, almost all assessors are qualified (level 4). If not they are undergoing training.	Yes, as the only qualified assessor in the organisation is the one normally conducting (i.e. verifying and ratifying) assessments. However the majority of supervisors and managers do not have an interest in assessment. The plan is to provide some basic training in assessment but it would not be possible to get them to undertake qualification as assessors.	All assessors are level 4 workplace assessor accredited and all have worked in the industry, so they have the competencies themselves. Also assessors have regular meetings for discussions and the co-ordinator sees students and conducts spot checks.

Interview summary cont.

Questions	Organisation					
Generalisability	P	Q	R	S	T	U
16. Can the outcomes for an individual assessee on an assessment task be affected by incentives (to either assessee or assessor), special forms or motivation, etc.	No	No	No	Not in CBA; the style of instruction should not make any difference. They are either going to learn it or they are not. If the assessment is transparent, incentives don't come into it.	Yes, motivation of the student has to be a factor. Assessors may also be influenced for a variety of reasons. However, the involvement of two people in the assessment (supervisor and qualified assessor) reduces this risk.	All students do the same assessment, and all assessors look for the same basic content. But, students may react differently to different lecturers and put more effort in and perform with more flair with one than they would with another.
17. Can you think of any examples where the results of an assessment have been used for purposes other than those for which they were designed? (e.g. an assessment of OHS actually being used as a test of knowledge of English)	No	No except that people with disadvantages don't get employed!	Yes, cited a case where a performance appraisal 'tool' was used to select a candidate for a job.	No, because, it doesn't come into the range of variables, nor into what evidence is collected. Competency standards are closely adhered to.	No	This would not happen. However, if spelling or grammar were bad, they would be asked to resubmit work with the errors corrected. At the higher levels, where trainees are writing procedures and processes, adequate spelling and grammar are expected.
18. Continuing this area, can you think of any assessment tasks where the outcomes may be influenced by factors not strictly relevant to the job, such as ethnicity, gender, socio-economic status, age?	No, they stay within guidelines.	Careful to follow legislative guidelines.	Believes the 'human factor' operates in the case of some assessors and sees monitoring as way of overcoming problem.	All students are treated the same regardless of gender or other factors. There are however special instances, such as physical disability and language and literacy difficulties.	No, language and literacy skills are a potential problem but the use of selection tests ensures such trainees are not employed.	Assessors are very careful about the language used and the environment the students are in. If there was a problem, it would be looked at in conjunction with the Institute's equal opportunities person. A different way of assessing may be used.

Interview summary cont.

Questions	Organisation					
Consequential	P	Q	R	S	T	U
19. Following on from the above, can you think of any cases where there are social consequences for the assessee, not related to the work-specific purpose of the assessment?	No	In other areas of company it has been observed that IT graduates from overseas lack English language skills and therefore are not 'team players'	No	No social consequences have been observed.	No, it has not happened, but the potential for it to happen exists (e.g. where trainees who fail to achieve competence may, against their will, not be retained as employees.)	There could be some social consequences which the lecturer/ assessor may not necessarily be aware of. There was one case, for example, of a student who couldn't cope with the assessments and withdrew from the course. The consequence maybe that the student has gone away with a lowered opinion of herself.

Relative importance of the eight types of validity evidence

The following table summarises the awareness of and actions taken by the participants in relation to the eight types of validity evidence. It shows that most respondents were aware of the issues defined by Nitko's separate identification of eight facets of validity. The types of validity not well covered are exactly those that might be expected to attract less attention, since they are related to the interaction between employment practices and the rest of the world, rather than to the core business of companies (for example, consequential and to a lesser extent generalisability). These facets would be more easily perceived by third-party observers.

Kind of evidence (In Nitko's terms)	Awareness and action amongst participants
Representativeness of content	Assessments were regarded as covering the training program adequately, and the assessment results were generally thought to reflect work practice . While it was generally agreed that the balance in the assessments matched the emphasis on the job, this was acknowledged as being a problem in some companies. In some cases it was indicated that assessments (and training) were not 'up to date'. In general, it was not thought that it would be possible to omit any of the present assessments. It was generally agreed that assessment tasks were worthwhile in themselves and actually fostered the learning process.

Relative importance of the eight types of validity evidence cont.

Kind of evidence (In Nitko's terms)	Awareness and action amongst participants
Internal coherence	Where evidence could be produced, it indicated that there was substantial consistency between results of different assessment tasks. However this emphasis on coherence was somewhat weakened by the varying approaches to procedures for handling the employee who did not succeed at a task at the first attempt. This is significant because a procedure which allows an indefinite number of repeats will 'appear' to generate consistent results when in fact the most likely underlying truth is that employees are performing differently on the various assessment tasks.
External consistency	In some cases subsequent workplace performance appraisals can generate evidence regarding consistency with job requirements, but nobody interviewed had a policy of checking on 'predictions' made by assessments. Only in a minority of cases were additional assessments used which would allow identification of 'outstanding' employees.
Reliability	Approaches to the issues arising with respect to reliability (or stability of assessment results) varied considerably amongst those interviewed. Most agreed that the time at which an assessment is carried out is important in the sense that there are times in a shift, or times of a day, when employees perform 'better' or 'worse' and it is important to 'standardise' the assessment time. (This in itself will mean that the assessment results and standards might not be achieved by employees at all times during the working day and also runs counter to the notion that assessors should use the 'range of variables' laid down in the standards.) Some believed that their standards are set in such detail that all assessors are forced to assess to the same standard; others believe that some form of moderation, even if only during the training of assessors, is necessary to ensure uniformity of standards. Companies are also divided on the usefulness of having 'back-up' tests to use when results of a test seem unreliable for any reason (as evidenced by erratic performance of those being assessed).
Substantive	Participants tended to see the capacity to deal with new situations as arising largely on 'the shop floor', and assessment of this is believed to be important. There was some opinion that 'rote learning' plays a role at certificate II level but less at higher levels.
Practicality	In general, participants were from companies or organisations in which assessment is a relatively high-impact cost, with most assessment being on a one-to-one basis. Cost-effectiveness of the assessment procedures was acknowledged as being an issue, but it is treated as incidental rather than a specific matter requiring attention. There are general efficiencies in that participants believed that qualifications were nationally recognised and their trained staff were sought after by others. In general, assessors performed well, but it was acknowledged that there was some room for improvement.
Generalisability	While those interviewed made it clear that formal (legislative) requirements with respect to potentially discriminatory practices were carefully observed, the idea of any form of favouritism in assessment, or misuse of assessment results (thus invalidating the generalisability of assessment results), was not easy for them to address. Perhaps this is to be expected, since those most likely to detect instances of 'misbehaviour' of this kind are third-party observers (who for example, might be researchers into social justice issues) and no interviews took place with third-party observers in this study. On the other hand, it is possible that while such cases occur they are so infrequent that those in the positions from which participants were chosen would not 'normally' be aware of such cases or the procedures which are followed.
Consequential	As was the case for some aspects of generalisability, this was not an area with which the participants were familiar. Once again, third-party observers are more likely to be aware of any effects in this area.

Key findings

The key findings of the study along with some suggestions for change and improvement are presented below.

Reluctance to participate

In any study of this kind some of the invited participants will decline, some relatively quickly, some after more extended consideration. In the VET sector in recent years, the participation rate amongst those invited to take part in research studies has generally been quite high.

To a limited extent that pattern has been followed in this study: most of those invited to participate agreed quickly and co-operated extensively with the researchers. However, this was not universally the case, and some private sector training and assessment organisations refused to participate. This placed some limitations on what could be achieved in the study, in part, because of the time consumed in (unsuccessful) negotiations, and in part, because of limited information being available from this group of potential participants.

This group may be termed the 'the training industry'. These potential participants were not *employers* of the persons who are being trained and assessed, they were training/assessing organisations whose core business lies with persons already employed.

As a result of this low level of participation by this group, the findings in this study are not comprehensive: we are able to say little about the validity of assessments conducted in such organisations. We do wish to express our concern, however, that these organisations, which are often relatively new players in the training and assessment of young Australians, were reluctant to allow any study of their assessment practices.

Process of auditing

Part of the process of continuous improvement or quality assurance of assessment is the conduct of a regular auditing process in which a third party verifies that assessments are being conducted in compliance with the formal requirements. In the present structure the State training authorities have this responsibility and the National Assessment Principles specify that assessments that are carried out 'will be subject to audit'.

Such an audit process would assist materially in establishing the validity of assessments being carried out. However, it is not clear that this process of auditing is well established or effective in all States and Territories, with a probable

consequential decline in confidence about assessment results and their use. Indeed, the recently completed report into the quality of vocational education and training in Australia (Senate Committee 2000) noted, in reference to submissions that had been received:

> *Other submissions indicate that there is a widespread view that state and territory authorities are less than thorough in administering the Australian Recognition Framework. Concerns and criticisms relate to the lack of consistency, rigour and integrity of processes for registration, performance monitoring and auditing of providers.* (p.122)

The extent to which it is possible to audit such a large system, beyond the consideration of compliance with procedures as opposed to a testing of those procedures, is also unclear. Although Principle 9 of the National Assessment Principles is that *Assessment systems must incorporate mechanisms for recording, storing and accessing assessment outcomes*, in the course of this study this appeared to be an area of some (by no means universal) weakness. While there is careful recording of results, storage and access seemed to be much less well developed, with the level of detail being, in some cases, quite minimal and less than would be needed for an audit. It is also important to note the current emphasis by State training authorities tends to be that of a 'front end' or compliance audit doing little more than checking an RTO's capacity to conduct and facilitate assessment. As the Senate Committee (2000) enquiry points out:

> *... whether the provider actually performs against the standards is not simply a question of having delivery and assessment skills and using them. It is also about whether persons who are assessed as competent against the standards are* actually competent. (p.121–2 [emphasis in original])

In other words, an audit of outcomes is also needed. The present study has not attempted such an audit, but the focus on validity has led to a seeking-out of the kind of data that would be required for an audit.

Use of 'RPL'

An evolution in understanding about RPL is currently occurring. Although there have been improvements since the days of 'considerable confusion' identified by Wilson and Lilly (1996), there is still not a universally agreed understanding about the use of RPL.

'Recognition of prior learning' may occur in two ways: either by examining existing documentation and thus determining that the holder of the documentation has previously acquired the skills included in the training program for which RPL is being assessed or, in the absence of such documentation, assessing in some other way that the applicant has previously acquired, and now possesses the relevant skills.

The first of these two ways is really a matter of recognising an assessment made by some other qualified person or institution, and giving the applicant credit for that earlier assessment. In other words, credit is transferred from the first person/institution to the second (they may in fact both be within the same institution), and this has accordingly been known as 'credit transfer'. The original

distinction between the two ways of recognising prior learning was to label them as 'credit transfer' and 'recognition of prior learning' (with the latter being intended to be reserved for cases involving a new assessment of skills).

In practice, however, the term 'recognition of prior learning' for a time became almost universal, with 'credit transfer' falling into disuse. One of the most common ways of conducting recognition of prior learning has nevertheless been through the use of documentation obtained from another person/institution, and this has led to the use of the term 'recognition of current competencies' to refer to a *current* assessment as opposed to the use of 'old credentials'. Kenyon et al. (1996) noted the 'strong support in the workplace for using the term recognition of current competency' and argued for its use in a broader context that included institutional training. Indeed, RCC continues to gain favour as the preferred approach since it concentrates on competencies that a trainee currently possesses rather than those once possessed. Employers sometimes find that a new employee with an RPL-based qualification needs refreshment training in order to 'get up to speed'.

In this study we have had to deal with both notions of recognising skills (a current assessment and an old assessment), and we will try to distinguish clearly between them in what follows.

There are several facets of validity involved in this consideration. Two of these are *generalisability* and *external consistency*. Both imply that an assessment result should be portable (in an appropriate context).

One issue that arises from this is the matter of company-specific requirements and the limitations these place upon the use of RPL/RCC. Any 'skills' related to company-specific requirements are not, in general, portable, and therefore should not generally be assessed by RPL/RCC. Furthermore, these are 'skills' that trainees can not expect to acquire in off-job training. This leads to differences between on-job and off-job training (for the same qualification), and these differences are more or less significant depending upon the relative importance of company-specific requirements.

The use of RPL/RCC is further limited (in this particular context) if holistic or integrated competency assessment is being used, but this is discussed as a separate issue.

A second issue is a general matter of non-comparability. RPL/RCC assessment results, by their nature, have not been achieved in the same circumstances as the assessment results for those assessed in the mainstream way. While every effort may be made to ensure that RPL/RCC results are comparable, the *reliability* of such assessments will always be doubtful. This is not because those carrying out RPL/RCC are in any way untrustworthy, but because there are no inexpensive ways of verifying the comparability of the results obtained in two different ways.

Our study showed that RPL/RCC is frequently a major component of the assessment of assessors, especially in workplace environments. Yet there is little recognition of the problem presented by including assessments of company-specific requirements outlined above. Assessors who have gained a qualification with a substantial component of company-specific RPL/RCC are unlikely to deal effectively with the *generalisability* and *external consistency* types of validity.

Relevance of facets of validity to CBA

Are Nitko's facets of validity relevant in a competency-based assessment environment? Isn't it the case that the trainee who demonstrates competence is actually simultaneously demonstrating that the assessment is valid?

Unfortunately it is not so. This is easily seen when we consider what the 'demonstration of competence' may mean. From our interviews, it is clear that a trainee may be allowed a number of attempts to demonstrate a competence (in some cases unlimited attempts may be made).

We can consider a simple situation: suppose a trainee was not able to demonstrate a competence at the first attempt, but was able to demonstrate it at a second attempt. What exactly do we learn from this information? Which of these scenarios is correct?

1. Does it mean that the trainee did not have the competence at the first attempt but undertook further training and thereafter did have the competence?
2. Does it mean that the trainee demonstrates the competence on about half the occasions it is required?
3. Does it mean that the trainee doesn't really have the competence at all, but had a lucky fluke at the second attempt?
4. Does it mean that the trainee had the competence right from the start but that the assessment (or the assessor) was defective in some way?
5. Does it mean that the trainee was assessed at 'the wrong time in a shift' on the first occasion but at 'the right time on a shift' on the second occasion?

We *do not and cannot* know which scenario is correct. What is more alarming, we could propose a similar set of scenarios about trainees who *succeed* at the first attempt. The reality, which is generally recognised in training packages and assessment methods, is that *only through multiple assessments* during which competence is *consistently* demonstrated can we choose between the scenarios outlined above.

The facets of validity that have been defined are useful in seeking additional evidence that the assessment outcome (the result of the assessment) will be the 'right' one, given a particular assessment result. As we demonstrate elsewhere, in general, those responsible for carrying out assessments are aware of the need for this additional evidence, and take steps to boost validity for the facets identified.

Workplace vs college assessment

Elsewhere we make brief reference to some of the differences between workplace and training college assessment (or on-job and off-job assessment). In this section we will focus upon one area that we see as presenting potential difficulties. This is the matter of integrated competency assessments (sometimes called holistic assessments).

Integrated competency assessments may be viewed as highly desirable developments; we do not expect trainees to demonstrate their competencies in isolation, or one at a time. (Imagine a car driver who could demonstrate only one

relevant competence at a time!) But while an ICA can be regarded as desirable, there are also complications and some of these emerged as issues of validity were being examined.

These complications can arise as follows:

- when an ICA is easier to conduct for some trainers/assessors than for others
- when an ICA generates performance evidence that conflicts with the assessments of individual units
- when an ICA generates performance evidence that matches exactly the evidence generated by assessment of individuals units

In the first case, where an ICA may, for example, be more easily fitted into the structure of a non-employer assessor, there will be advantage or disadvantage for trainees depending on the circumstances of their assessment (lowering the validity of the overall assessment package). In the second and third cases, a conflict implies lowered validity of one or the other of the assessments (the *reliability* facet), while an exact match (the third case) is inefficient because it provides no new information (*practicality* facet of validity). Some slight difference between the two approaches may be useful—but may raise doubts in the minds of users.

Another doubt can be found in the structure of the ICAs. When constructing an ICA, the assumption is made that the whole is greater than the sum of its parts. That is, the combined assessment task covers all the competencies identified in the individual units. Examination of some examples raised questions as to whether or not this aim was being achieved—a question of the adequacy of *content* evidence.

There is a second difference between on-job and off-job assessment that we want to note in passing: for on-job assessment the locus of responsibility is very narrow—the employer lives with the assessment *outcomes*. For off-job assessment on the other hand, responsibility is rather more difficult to locate (since the assessor produces assessment *results*, but does *not* have to live with whatever use is made of the results—the assessment outcomes). The joint responsibility of training colleges and employers in correctly using the results of off-job assessment is a significant issue when making decisions about validity.

Graded assessment vs CBA

The issue of the relationship between graded assessment and competency-based assessment is a complex one. It may be argued, from a theoretical standpoint, that there is no place for graded assessment in a competency-based assessment system. If the purpose of an assessment were solely to separate those who have shown that they are competent from those who have not shown that they are competent then a graded assessment would make no contribution to the winnowing process.

Does the intent to specify an assessment outcome in terms of competency have any effect on the way in which an assessment result is specified? On the one hand, we have to require that the assessment itself fits in with the general framework of being competency-based. But on the other, we ought not to be artificially constrained as to how that assessment is conducted or a result obtained. So are there circumstances in which an assessment that leads to a graded result is appropriate, even within a competency-based framework?

What is it that might justify a graded result? First, there is an assertion that performance in the given area is found to different extents in the relevant population. Second, there is an assertion that a system exists that can measure (with appropriate accuracy) the extent to which an individual in the population can perform.

From the purist viewpoint, this pair of assertions is not consistent with the notion of competence. The 'different extents' in the previous paragraph are just two in number, 'competent' and 'not competent', and the purpose of a competency-based assessment is to make that simple distinction.

The context within which assessment is carried out in Australia does not, however, meet this purist standard.

Consider, for example, some possible assessment outcomes in the retail sector: a person being trained may receive (for example) a Certificate I in Retail Operations, or a Certificate II in Retail Operations, or a Certificate III in Retail Operations, or a Certificate IV in Retail Operations. These four certificates form a sequence, one building on the other. In some companies this sequence is used to assign recipients to points on a pay scale. Not only are these certificates 'grades' of one form of competence (retail operations), but the language used in defining modules within them reflects the notion of an ascending sequence within a more narrowly defined competency: 'This unit builds on …' is one of the most common phrases found in the standards.

Furthermore, in many cases the performance criteria do not generate simple yes/no answers ('Communication in the workplace' is a convenient example). Apart from the inherent difficulty of making judgments about some communication skills, variation in interpretation by different assessors is also to be expected.

In the structure of these training packages, and in the consequences of an assessment, the notion of sequence and grading predominates. It would be artificial and inconsistent to assert that the notion of grading *automatically* has no place in the relatively small assessment component, given its implied use elsewhere.

This is *not* an assertion that *any* form of graded assessment result is compatible with a competency-based system that relies ultimately upon a statement of competence. Assessments must still be carefully and appropriately designed. But it is important to emphasise that a graded assessment result can be wholly compatible with a competency-based approach, and that indiscriminate rejection of graded assessment is not defensible.

Attitudinal factors

It is important in a study of this kind to ask whether or not there are 'other factors' 'unstated factors' which are relevant and which might influence assessment outcomes and so raise questions about the 'validity' of the assessments. Nitko's types of validity evidence do not address this issue directly, identifying instead factors which are 'irrelevant' but which might interact with an assessment outcome. Nevertheless, the issue of relevant but unstated influential factors ought to be addressed in this study, as in one way they challenge the current underlying training philosophy.

Are *all* those who are declared by their certification to be equally 'competent', (that is, equally useful on the job), equally likely to succeed? We do not think so.

There are other factors and these factors are the (usually) unspoken attitudinal ones—'bright-eyed and bushy-tailed'. Such factors contribute to employability, and to on-job success and are among the first sought by employers, but rarely appear in definitions of competence. To the extent that such factors are relevant, but unstated in training 'requirements', the national training packages are failing to deliver what industry needs.

In addition, where such factors are influential, they will reflect on the validity of assessments, especially, in Nitko's terms, in the facet of *external consistency*. This is not an area in which we have answers, but our consideration of a broad definition and understanding of 'validity' leads us to challenge the completeness of an assessment package which fails to address a central expectation that employers have about those declared to be 'competent'.

Practical help: A diagnostic tool for improving the validity of assessment

Another key finding was that assessors clearly needed assistance in getting to grips with the greater complexity that comes with the new approach to validity.

As part of the project the researchers developed and piloted a diagnostic tool aimed at helping practitioners improve the validity of their assessments.

This is provided as appendix A of the report.

References

AERA (American Education Research Association) 1999, *Standards for educational and psychological testing*, AERA, Washington DC.

Cropley, M 1995, 'Validity' in *Key aspects of competency-based assessment*, ed. W Hall, NCVER, Adelaide.

Ebel, RL 1965, *Measuring educational achievement*, Prentice Hall, Englewood Cliffs NJ.

Gillis, S & Bateman, A 1999, *Assessing in VET: Issues of validity and reliability*, Review of research, NCVER, Adelaide.

Hager, P, Athanasou, J & Gonczi, A 1994, *Assessment: Technical manual*, AGPS, Canberra.

Kenyon, R, Saunders, J & Gibb, J 1996, *Improving RPL—A training providers' perspective*, NCVER, Adelaide.

Lindquist, EF, 1942, *A first course in statistics*, Houghton Mifflin, Boston.

Messick, S 1989, 'Validity' in *Educational measurement*, 3rd edn, ed. RL Linn, Onyx Press, Phoenix AZ.

NAWTB (National Assessors' and Workplace Trainers' Body) 1999, Training package for assessment and workplace training (BSZ98), ANTA, Melbourne.

Nitko, AL 1996, *Educational assessment of standards*, 2nd edn, Prentice Hall/Merrill Education Englewood Cliffs NJ, pp.44–5.

NTB (National Training Board) 1992, National competency standards policy and guidelines, 2nd edn, NTB, Canberra.

Senate Committee (Senate Employment, Workplace Relations, Small Business and Education References Committee) 2000, *Aspiring to excellence: Report into the quality of vocational education and training in Australia: Overview and recommendations*, Canberra.

Toohey, S, Ryan, G, McLean, J, & Hughes, C 1995, 'Assessing competency-based education and training: A literature review', *Australian and New Zealand Journal of Vocational Education Research*, vol.3, no.2, pp.86–117.

Wilson, J & Lilly, M 1996, *Recognition of prior learning*, NCVER, Adelaide.

Appendix A

Diagnostic tool for assessors

Guidance material for assessors that aims to improve the quality of the procedures by improving the validity of assessment is one significant outcome of the project.

IT IS IMPORTANT TO NOTE THAT THE APPROACH TO VALIDITY THAT FOLLOWS DIFFERS FROM THAT CURRENTLY USED IN THE TRAINING PACKAGES.

A guide for improving assessments

Improving your competency-based assessment skills by improving the validity of your procedures.

What is validity?

Validity is the extent to which the interpretation and use of an assessment outcome can be supported by evidence.

Validity is what provides assessments with quality and is concerned with the soundness of the interpretations and uses you make of your assessment results.

Validity is not a property of the assessment instruments themselves.

What should you do to improve the validity of your assessment procedures?

Before this question can be answered we need to know which, if any, aspects of your assessments need attention. As there are a number of possibilities to consider, we have framed a series of questions to help you identify any problems you might be having with validity.

Read the questions and answer YES, NO or NOT SURE to each.

When you answer NO or NOT SURE you are directed to another section of the document where you will find additional information to clarify and hopefully help you deal with any problems you are having.

Note, in what follows we have used the term 'trainee' to include 'student' and 'apprentice'.

13 questions about your assessment procedures

1. Are you sure all performance criteria in the industry standards are matched by tasks and questions in your assessment instruments and that no important or essential criteria have been overlooked?

☐	Yes	
☐	No	] Go to pp.45–46] *content*
☐	Not sure	] evidence

2. Are you sure that your assessment instruments, taken as a whole, cover all the following components?
   - task skills (performing to an acceptable level)
   - task management skills (managing a number of different tasks within the job)
   - contingency management skills (responding appropriately to unexpected events)
   - job/role environment skills (working with others/working as part of a team)
   - transfer skills (using skills learnt in one situation in a new or different context)

☐	Yes	
☐	No	] Go to pp.45–46] *content*
☐	Not sure	] evidence

3. Are you confident your assessment procedures are cost-effective?

☐	Yes	
☐	No	] Go to p.48] *practicality*
☐	Not sure	] evidence

4. Are you sure you have the necessary skills to be a competent assessor?

☐	Yes	
☐	No	] Go to p.48] *practicality*
☐	Not sure	] evidence

5. Are the results of the various components of your assessments (written, oral, practical, workplace) consistent with one another?

☐	Yes	
☐	No	] Go to p.48] *internal*
☐	Not sure	] *coherence*] evidence

6. Is there a clear relationship between the trainees' outcomes on your assessments and their performance at work or in further learning they undertake?

☐	Yes	
☐	No	] Go to p.49] *external*] *consistency*] evidence
☐	Not sure	

7. Are the outcomes of your assessments consistent with evidence from other sources?

☐	Yes	
☐	No	] Go to p.49] *external*] *consistency*] evidence
☐	Not sure	

8. Are you confident that other assessors working in your area would independently come to the same conclusions about a trainee's performance as you?

☐	Yes	
☐	No	] Go to p.50] *reliability*] evidence
☐	Not sure	

9. Are you confident that, if you assessed the same performance by a trainee on two occasions, one week apart and without further training, you would get the same result?

☐	Yes	
☐	No	] Go to p.50] *reliability*] evidence
☐	Not sure	

10. Are you confident that considerations of a trainee's sex, ethnicity or socio-economic status do not influence the way you use the results of your assessment?

☐	Yes	
☐	No	] Go to p.50] *generalisability*] evidence
☐	Not sure	

11. Are you confident that the way you use the results of your assessments does not lead to inappropriate social consequences for the trainees?

☐	Yes	
☐	No	] Go to p.51] *consequential*] evidence
☐	Not sure	

12. Do your assessment tasks include the application of thinking skills other than rote learning? (e.g. are you sure your tests require the trainees to apply critical thinking to solving problems?)

☐	Yes	
☐	No	] Go to p.51] *substantive*
☐	Not sure	] evidence

13. (Only for assessors who use graded assessments) Do your grades predict the performance of your trainees at work or in further learning they undertake?

☐	Yes	
☐	No	] Go to p.49] *external*
☐	Not sure	] *consistency*] evidence

What you need to know and do

About content

Matching standards with assessments (Question 1)

Checking assessment tasks and questions for evidence that content has to be dealt with involves a number of steps.

Most obviously, you can perform a cross-check between the *performance criteria* in the industry standards and your assessment instruments to make sure nothing has been overlooked.

However, the *performance criteria* themselves often require further interpretation. Consider, for example, the criterion *Evidence gathering activities are planned to provide sufficient, reliable, valid and fair evidence of competency in accordance with the assessment procedure'* (from the unit 'Conduct assessment' of the Training Package for Assessment and Workplace Training). Interpreting this criterion involves answers to a series of questions such as:

> What sort of evidence gathering activities?
>
> What's reliable, valid and fair evidence?
>
> What's sufficient?

For these answers you need to go to the *evidence guide* and *range of variables* section of the industry standard. When you do this, remember that the training packages have been written to cater for a wide audience, whereas you are, in all probability, wanting guidance specific to your own needs. This means you should interpret the standards with your needs in mind and not worry about the wider audience. For example, in the 'Review assessment unit' of the same package, the standard states:

> *Assessment methods may include a combination of:*
>
> - *work samples and simulations*
> - *direct observation of performance, products, practical tasks, project and simulation exercises*
> - *questioning*
> - *consideration of third-party reports and authenticated prior achievements*
> - *written, oral or computer managed questioning*

But if your assessment procedures involve only simulations with written and oral questioning, then you do not need to go beyond these in any review you carry out.

Another thing to avoid when developing assessment instruments is 'taking the easy way out'.

Some performance criteria are easier to assess than others. For example, it is easier to assess whether a trainee has satisfactorily cleaned up after doing a job than assessing the trainee actually doing the job. It is also easier to assess a performance that occurs repeatedly than one that occurs only occasionally at irregular times.

The trap you can fall into is to develop an assessment instrument that mostly consists of frequently observable, easy to assess items. If this means that essential skills are being overlooked, the result will be a poor (and invalid) assessment.

Most training packages identify the essential evidence and assessors need to check that their assessments are meeting these requirements. For example, any assessor who fails to include the occupational health and safety performance criteria identified in the standards has most likely produced an invalid assessment.

It is also important to ensure that your assessments are 'up to date'. This means checking that any skills and/or knowledge recently introduced into the training program have been included and any material no longer relevant has been deleted from the assessments. (It is not unusual for the people who have prepared the standards and the training program to take a while to act on the changes happening around them. Assessors need to point the changes out to those responsible and they should never assess anything that is irrelevant.)

Computer programming courses are examples of where new material must be included (and assessed) on a regular basis with a corresponding deletion of assessments related to material that is no longer relevant.

As obvious as this may seem, it is disappointing to observe how much irrelevant material can be found in some assessments. Although nobody uses Morse code any more, a recent survey of seafaring courses around the world found that it was still assessed as an essential skill in a number of countries.

Components (Question 2)

You should check that your assessment procedures, when considered as a whole, cover all the components that make a person competent namely:

- task skills *(performing at an acceptable level of skill)*
- task management skills *(managing a number of different tasks within the job)*
- contingency management skills *(responding and reacting appropriately to unexpected problems, changes in routine and breakdowns)*
- job/role environment skills *(fulfilling the responsibilities and expectations of the workplace)*
- transfer skills *(transferring skills and knowledge to new situations and contexts)*

In the early days of competency-based assessment the most frequent criticism was that it was only about the technical skills in doing a job. The analogy of 'trained monkeys' was frequently used. Things are better today, although technical or task skills remain the dominant form of assessment in many industries.

When developing your assessment procedures you need to analyse the requirements of the job to which they relate and make sure all have been covered.

By way of example, consider the requirements of a receptionist's job as they relate to the components.

Component skills	Competent receptionist
Task skills	Meets company standards for: • word processing speed and accuracy • telephone answering skills • customer service skills • clerical skills
Task management skills	Demonstrates ability to manage tasks so that needs of management, other staff and clients are all met to their satisfaction within an acceptable time
Contingency management skills	Demonstrates ability to reschedule work to handle unexpected problems such as power failures, sudden illness of staff etc., so that there is minimum disruption to the normal work flow
Job/role environment skills	Demonstrates ability to work well as part of the 'front office' team and to have good interpersonal skills with peers, managers and customers
Transfer skills	Demonstrates the ability to transfer skills learnt in dealing with problems presented by one customer when similar problems occur with another customer

About practicality

Cost-effectiveness (Question 3)

Stripped of its numerous economic and social elements, a cost-effective performance assessment is a quality assessment carried out for less cost than other assessments of the same performance.

Assessors who want to investigate the cost-effectiveness of their assessments need to begin with the understanding that the one-to-one assessments which are a predominant feature of a competency-based system must inevitably be *more costly* than most other forms. The issue is, therefore, whether or not the outcome is *more effective* than the alternatives. To make this decision an assessor will usually need to look beyond the assessment procedures alone. For example, if the introduction of competency-based assessment increased the cost of a training program from $1000 to $1250 per month, but (because of increased effectiveness) reduced the training time from six months to four months, this would be a clear demonstration of a more cost-effective assessment, as the overall cost would be reduced from $6000 to $5000.

Unfortunately, most decisions about cost-effectiveness require the assessor to take into consideration a range of factors. Furthermore, in most cases, monetary costs are only part of the consideration. To help assessors reaching a decision about cost-effectiveness the following guidelines should be applied.

- The more varied the assessment evidence you gather, the more likely you are to make the correct judgment about competence.
- All essential skills, in particular occupational health and safety skills, must be assessed.
- When assessing the performances of a group of trainees at the same time, the larger the group, the more likely you are to make an incorrect judgment.
- If the choice is between cost and fairness you must choose fairness.
- Frequent and repeated assessments are more likely to result in a correct judgment than a single end-of-course assessment.

- If you are thinking about changing to a different method of assessment but cannot say how the new method will add value to the training system then you probably should not change.

Competent assessors (Question 4)

Dealing from day to day with a system based on defined standards of performance can sometimes have assessors worrying that they are often making judgments from limited, even dubious evidence. And even assessors with the currently required certificate IV or diploma qualification might feel that they are not sufficiently well prepared for their task.

Apart from pointing out that 'you are not alone' and that many assessors worry about their state of preparedness, we offer the following guide:

- There is no substitute for subject expertise. If you don't have skills in the area you are assessing, then you should either acquire those skills or enlist the help of someone who has the skills.
- It is unlikely that your assessment skills will improve in isolation. Discuss your assessments with a knowledgeable colleague or better still, join an assessor group (or start one if there isn't one in your area). 'Moderation' groups like those suggested in 'About reliability' (p.49) are examples.

About internal coherence

(Question 5)

Getting the same or very similar results from the different assessment instruments you use can give you confidence that you are using appropriate methods, but there can be a downside to such results.

Indeed, when you find your trainees' performances on written, oral, practical and workplace tests show little or no variation, you should take a closer look at the tests. It might be that you are testing the same, or a narrow range of skills with each form of the test and that, overall, your assessments are not giving the necessary coverage.

So the things to remember are:

- People vary in the ability to perform on different types of assessment (for example theory and practical), therefore some variation in results is to be expected.
- Although you are not seeking identical results on each test you are seeking similar results, any test that produces results markedly different from the others needs to be checked.
- Don't fall into the trap of dropping all tests bar one because they give the same or similar results. Such an action inevitably means only a limited amount of evidence can be gathered. A single test is unlikely to reflect work practices in any meaningful way.

About external consistency

Relationship between assessment and performance (Question 6)

For an assessor working in a company and assessing trainees or apprentices there isn't much external consistency to consider. One of the great strengths of the competency-based approach in workplaces is that the work people do is not external but, is integral to the assessment process. Therefore, workplace assessors can only judge 'external' consistency by looking for evidence of a relationship between the results of their assessments and some future event. For example, do trainees who under-perform on assessments leave the company or move to other kinds of work?

Consistency of assessment (Question 7)

The key to confirming that your assessments are on the right track in this area is to keep good records and follow up ex-trainees over time. If you find that someone you deemed to be 'not yet competent' five years ago is now your boss then it just might suggest your assessments lack external consistency!

The above points also apply to assessors working in training colleges or as individual consultants. However, if you belong to either of these categories you have a rich source of information to tap. By going to the companies now employing your ex-trainees you can ascertain what the relationship is between the results the trainees obtained and their subsequent performance at work.

Grading (Question 13)

For those assessors who also assign grades of competence, the test of their grading procedures is done in a similar way. The grades awarded to trainees should be reflected in their comparative performances after the assessment. If this is not happening then your grading procedures need revising.

About reliability

Reliability of different assessors (Question 8)

Assessors face the possibility of making an error every time they make an assessment. These errors can occur for several reasons. A major source of error when more than one assessor is involved relates to their different interpretations of the competency standards.

Consider, for example, the Communication skills performance criterion 'messages or information recorded and passed on promptly'. If one assessor interprets 'promptly' as within 10 minutes and another interprets it as within an hour, then a trainee who consistently passes on messages in around 30 minutes will be deemed competent by only one of the two assessors.

This interpretation problem occurs throughout the training packages and is a substantial issue for assessors, particularly those working in isolation. However, if two or more assessors are working together, they can meet regularly to discuss trainee performances and reach a consensus on the standard for their RTO. Better still they should meet with assessor colleagues from other RTOs and industry and

invite industry representatives to give their view on what constitutes *meeting the standard required in employment*. Although there are benefits in meeting face to face, this may sometimes be too costly. When this is the case, alternatives such as teleconferencing, faxes and email can be used effectively.

The process undertaken at these meetings is known as *consensus moderation* and this can occur at a number of levels, for example, RTO, regional or State/Territory. It is also important to appreciate that the amount of consensus resulting from these meetings is dependent on who participates. For example, if the participants all come from the one TAFE institute then consensus is confined to that institute, nobody can be sure that assessors from another institute or a workplace would arrive at a similar consensus. Therefore it follows that the greater the range of backgrounds of the participants attending the meetings, the broader the consensus and the greater the likelihood of quality evidence of reliability.

Any assessors who are unsure whether or not their assessments are consistent with those of their colleagues need to become part of a consensus moderation process. Indeed, the staff development activities that are part and parcel of these meetings benefit all assessors and cannot be too highly recommended.

Reliability of individual assessors (Question 9)

As well as being consistent with other assessors you also need to be sure that you behave consistently. Assessors have good days and bad days like everyone else in the workplace—the challenge is to make sure your judgment when making an assessment is not affected.

Essentially, this is something only you can control. You need to recognise the signs—excessive tiredness, illness or perhaps a recent argument with the person you are about to assess. Just as you are looking for consistency in the performance of your trainees, so too are they expecting you to be consistent.

If there are reasonable grounds to believe your performance as an assessor will differ substantially from your normal performance then you should postpone the assessment.

About generalisability

(Question 10)

We want our assessments to apply equally to any trainee who decides to undertake them. We want them to apply to the trainees in general, irrespective of sex, social class or ethnic background. And we don't want any personal prejudices we may have, for example, any personal antagonism we have towards fans of the Bombers or the Broncos must not affect our assessment of a trainee who happens to be a supporter of these clubs.

As obvious as this may seem, it is still possible to cite cases where assessments fail the generalisability test. For example, females being banned from taking an assessment because they were said not to be strong enough to be scaffolders and mildly colour blind individuals being prevented from becoming electricians.

Assessors need to be aware of the threat to validity that comes with failing to ensure their assessments are generalisable.

About consequences

(Question 11)

Assessors in the VET area have not concerned themselves greatly with the consequences or social impact of their assessments. Some of their colleagues who assess students in Year 12 show a much greater awareness of the issue. For example, it is possible to argue that one consequence of the highly selective Year 12 examination is the negative impact it has on learning. Some teachers at this level complain that all they do is train students to answer possible exam questions.

It is possible to see how negative impacts can also occur in the VET area. For example, if an assessor regularly fails to assess (or assesses imperfectly) a critically important OH&S skill, then trainees who 'graduate' from a program where this occurred can become dangers to both themselves and their fellow workers.

As an assessor you need to be aware that there are possible side effects or consequences your assessment can have that you never intended. Any negative consequences need to be addressed and this might require changes to your assessment procedures.

About substantive evidence

(Question 12)

There is nothing wrong with rote learning, indeed remembering a frequently used phone number saves a person much time and effort. However, only assessing tasks that can be readily learnt by rote means we miss the vitally important critical thinking and problem solving skills that are a part of being competent.

But because we cannot see people 'thinking' or 'understanding' we have to infer these skills from their performances.

Assessors therefore need to design assessments that tap into the areas of critical thinking and problem solving. If the assessment task involves a practical process (e.g. preparing a meal) or a written assignment (for example, evaluating a diet) then, in addition to observing and judging the performance, additional evidence should be gathered through asking 'What ... if ....' questions such as:

- What would you have done if you discovered the oven was set 50°C too low about half an hour after you began cooking the meal?
- What changes should be made to the (specified) diet if it were to be used on children younger than 12?

These questions can be asked orally and therefore do not involve large investments of time.

It is also important to appreciate that there is usually no one right answer to these sorts of questions. Assessors need to use their experience to judge what is acceptable.

Appendix B

Case studies

Woolworths Supermarkets
McDonald's Australia
National Pharmacies
DJ Foreman & Associates
Logistics Training Unit, Douglas Mawson Institute of TAFE
Retail Training, School of Business Studies, Regency Institute of TAFE

Woolworths Supermarkets

Background

The investigations of Woolworths Supermarket Limited centred on the company's certificate II level qualification. Certificate II is the minimum entry-level qualification for the retail industry. A range of certificate II assessment instruments related to various aspects of customer service were the focus of the study.

Woolworths Supermarket Limited is a registered training organisation that conducts all its training and assessment in house.

The information on the program and the approach taken to assessment was supplied by the national training manager and the national project manager—traineeships, both of whom were qualified assessors (certificate IV Level).

The Woolworths training program has been developed by the company and has its own training manual which is aligned to the National Wholesale Retail and Personal Services (National WRAPS) Industry Training Councils' training package.

Probably the most notable feature of the Woolworths approach to assessment is the emphasis placed on quality through the Woolworths Training Quality Management System (WTQMS). There is a WTQMS co-ordinator in each State and Territory and a national WTQMS manager.

The WTQMS group meet by video conference once a month and assessment issues tend to dominate the agenda. One matter currently under consideration by the group is how best they can ensure there is consistency in the assessment of standards between the States and Territories. WTQMS has addressed the consistency issue by arranging for the States to audit each other's procedures. WTQMS co-ordinators from one State travel to another to do the audit and their procedures are, in turn, audited by a co-ordinator from another State.

There is also a proposal for each of the WTQMS co-ordinators to observe their assessors conducting an assessment at least once every 12 months. This should help deliver better intra-state consistency. Consideration is also being given to the possibility of exchanging work samples and video-tapes of employees performing various tasks between assessors in the different States and Territories so that a common consensus can be reached over what constitutes 'competent' and 'not yet competent' performance for Woolworths. This process (a form of consensus moderation) should make a substantial contribution to the company's desire for national consistency.

Woolworths standards and assessments

Part of the assessment review carried out with each RTO in the study involved a cross-check of the RTO standards against the assessment instruments. The purpose of this part of the review was to gain answers to the questions as set out below:

1. Are all the RTO's *performance criteria* covered by the assessments?

   The Woolworths program included a self-assessment exercise in which each trainee was required to check off the fact that they had achieved all the

performance criteria in each module of training. The final assessment conducted by the assessor, however, covers only a sample of these criteria.

2. Do the assessments cover the different situations and equipment specified in the *range of variables?*

   Woolworths organise its range of variables statements differently from those of national WRAPS. The Woolworth's statements are subdivided into four sections:

   - general context
   - worksite environment
   - sources of information/documentation
   - applicable State/Territory regulations and requirements

   This organisation makes for easier reading than the national materials and enables a more customised or focussed approach to training and assessment. Nevertheless, there is a formidable range of contexts, documentation and equipment described in each section and only a sample is covered in the assessment instruments.

3. Are the essential requirements and the underpinning knowledge, skills and attitudes as specified in the *evidence guide* covered in the assessments?

   In a similar presentation to that used by National WRAPS, the evidence guide begins with a list of the *Critical aspects of evidence to be considered*, an example of a critical aspect is 'the ability to anticipate products and services required by customers'. Other critical aspects relate to the ability to transfer skills to new contexts and consistency of performance.

   Underpinning knowledge and skills evidence requirements are also listed along with suggested techniques for the assessor to use when gathering the evidence.

   The assessment instruments that have been prepared to judge whether or not a performance such as 'the ability to anticipate products and services required by customers' has been achieved, is for the most part, simply requiring the assessor to make a judgement based on observations over time. In other words with some criteria the matching of the assessment instrument to the performance is of minor importance. What counts most is the accuracy and consistency of the individual assessor.

   However, assessing underpinning knowledge and skills such as 'knowledge of store policies and procedures' can be more closely related to the quality of the questions on the assessment instrument.

   Woolworths clearly appreciates this distinction with its reliance on formal assessment of underpinning knowledge and skills on the one hand, and the efforts to improve the quality of assessor performance (through the WTQMS) on the other.

4. Are the *key competencies* assessed separately or is the acquisition of key competency assumed in some way?

   The Woolworths training program follows the practice of National WRAPS in listing the key competencies:

- collecting, analysing and organising information
- communicating ideas and information
- planning and organising activities
- working with others in a team
- using mathematical ideas and techniques
- solving problems
- using technology

Each key competency can be assessed at one of three levels with level 1 the 'basic' level. The company's certificate II training manuals simply identify each of the key competencies as being covered in the program at the basic level. Essentially this means undertaking activities efficiently and with sufficient self-management to meet the requirements for competency in the activities. There are no assessment instruments specifically devoted to assessing key competencies, although it was noted that the final assessment for the certificate II qualification contained a number of problem-solving exercises.

Conformity with retail industry guidelines

As noted above, the Woolworths supermarkets training and assessment manuals have been aligned with the National WRAPS Training Package. Registered training organisations such as Woolworths are able to customise their standards and qualifications provided they remain consistent with industry requirements.

The Woolworths training and assessment manuals were cross-checked against the National Retail Training Package assessment guidelines. These require an RTO to meet a range of criteria related to such things as assessor qualifications, quality assurance and audit procedures, recording and reporting of assessment outcomes, the designing and conduct of assessments, appeals processes, recognition of current competencies, authentication and 'equal opportunity' legislation.

Assessor qualifications

Much of the assessment in certificate II is done in the workplace by supervisors who have been trained as assessors through a Woolworths program based on the Training Package for Assessment and Workplace Training.

Quality assurance and external audit procedures

The quality assurance and internal auditing of the assessment procedures are addressed by the Woolworths Training Quality Management System as outlined above in the background section.

External auditing is a more complex procedure as the processes are developed and managed by the State and Territory training authorities. Therefore, Woolworths must deal with eight separate bodies in order to fulfil the audit requirements as these differ around the country.

Recording and reporting assessments

Woolworths keeps records of their trainees on a State and Territory basis. There is currently no national database.

Design and conduct of assessments

The National WRAPS Training Package recommends two different types of assessment:

- *Integrated competency assessment:* where assessment for the Certificate II in Retail Operations is based on an integration of workplace competencies into an holistic activity. The units of competency are grouped into phases of interrelated units to facilitate the assessment process.
- *Unit assessment:* where assessment against individual units of competency is based on an integration of the performance criteria into an holistic activity for that unit.

  For certificate II the unit assessment is also seen as a means of measuring progress and may be carried out by the workplace coach/supervisor after which the integrated competency assessment becomes the formal assessment.

However, although it is under consideration, Woolworths does not currently use integrated competency assessment, preferring, for the present, to assess each unit separately. Other companies prefer not to use the integrated competency assessment approach and the issues surrounding this decision are explored more thoroughly in the *Key findings* of the report.

Appeals procedure

Woolworths have a mechanism in place which allows a trainee to appeal against an assessor's decision. The appeal process involving the trainee and assessor is mediated by managers in the training departments in each State and Territory.

Recognition of current competencies/prior learning

Woolworths recognises that individual trainees may already possess relevant competencies which they gained independently of the training program. Their policy is for trainees claiming to possess a competency to take the same assessments at the normal trainee intake.

Authentication of work

Assessment in the Woolworths program includes judgement on assignments and portfolios that are done outside normal work time and hence unsupervised.

When the products of unsupervised work are presented for assessment, the assessor is not usually required to authenticate that the work was done by the trainee. At this point the company does not see the need to authenticate unsupervised work preferring to trust the trainee will behave in an ethical way.

Equal opportunity legislation

Woolworths goes to considerable lengths to ensure that it fully complies with federal and State legislation regarding equal opportunity for disadvantaged groups.

Components of an assessment

A sample of the Woolworths certificate II assessment instruments were examined to see how well they covered the essential components of competency specified in the Training Package for Assessment and Workplace Training.

These components are:

- task skills (performing at an acceptable level)
- task management skills (managing a number of different tasks that make up a job)
- contingency management skills (dealing with the unexpected)
- job/role environment skills (working with others, working as part of a team)
- transfer skills (using skills learnt in one situation in a new or different context)

Woolworths' emphasis on workplace assessment makes it possible to observe the trainees' level of performance of the first four of these components over time. The assessor checklists give an adequate coverage of skills required to perform tasks, manage time, deal with the unexpected and work in a team. There are also written and oral tests which can be used to confirm the assessor's observations.

The component not covered well is *Transfer skills*. Perhaps this is not that surprising since the limited time of the Certificate II program makes it difficult to expose the trainees to new contexts. However, given the importance of skill transfer to employees and the organisation as a whole, it should be possible to assess this component when trainees move to take on new roles in the company.

McDonald's Australia

Background

The investigations of McDonald's (Australia) included an overview of training and assessment procedures in general, with a concentration on assessment for the Certificate II in Food Retail—McDonald's.

McDonald's (Australia) has been recognised as a registered training organisation by all State and Territory training authorities. Although extensive use is made of training videos and training manuals, the predominant focus for its training and assessment is the workplace, with refreshment and reinforcement sessions done on completion of the various courses at the company training centres.

Information on the McDonald's program and the approaches to assessment were supplied by company training consultants in Melbourne and Brisbane, both of whom are qualified assessors.

McDonald's do not use the terminology of the National WRAPS standards, preferring instead to keep to 'training modules', 'learning outcomes', 'objectives' and 'assessment activities'. In order to achieve national recognition of the company-specific materials, it was necessary to include some broader industry skills in a number of the McDonald's training packages. The company has evidence to show that the McDonald's certificates resulting from this process are highly regarded by other employers across the food service industry.

McDonald's is also noteworthy because of the age of its employees. The bulk of the 'crew'—the workers serving and preparing the food— are around 16 to 18 years. And promotion can be speedy for those that seek it. Crew trainers of 18 and crew managers of 19 are not unusual.

The assessments related to skill acquisition are covered in more detail in the following section; however, it is important to appreciate that McDonald's also uses a form of graded assessments called performance ratings with its employees. These overlay the other assessments and provide four grades:

- needs improvement
- good (meets standards)
- excellent
- outstanding

Performance review checklists are used to determine these grades. A performance review is frequently conducted when a trainee appears to be in difficulties with the training program and serves as a basis for counselling the trainee on how to overcome the difficulties.

McDonald's standards and assessments

As McDonald's does not use the national industry standards laid down in the National Wholesale Retail and Personal Services (National WRAPS) Industry Training Council Training Package, the review had to take a different approach from that adopted with other RTOs.

It was possible to confirm a match between the industry units of competence and McDonald's training modules. For example, there was a close parallel between National WRAPS Unit WRRCS.1A—*Communication in the workplace* and McD02—McDonald's module on workplace communication. The National WRAPS elements of competency for this unit each contained between two and six performance criteria whereas the McDonald's performance criteria for the learning outcomes needed to be deduced from the *assessment activities* that accompany each outcome.

A closer examination of McDonald's assessment activities showed the advantages that company RTOs have over other RTOs that must cater for a diverse clientele. Consider, for example, the performance criteria for the National WRAPS element of competency CS1.1 'Establish contact with customers', these are:

- *welcoming customer environment maintained*
- *customer greeted warmly according to store procedures*
- *effective service environment created through verbal and non-verbal presentation according to store policy*
- *questioning and active listening used to determine customer needs*
- *confidentiality and tact demonstrated*

McDonald's training and assessment covers each of these criteria but adds the store specific requirements that cannot be provided in the national standards. McDonald's trainees are *trained specifically* on how to greet customers, what the requirements of effective service are, what questions to ask, how to handle payments and deal with complaints. Furthermore, the performance of each of these skills is assessed on the job using the station observation checklists (SOCs) that include a specified time within which the set of performance criteria must be completed.

The SOCs are interesting instruments in that they serve both a training and assessment function. Training is done by a crew trainer taking the trainee through a process designed to have the trainee achieve the standard required to be competent in each of the skills on the checklist. The crew trainer then formally assesses the trainee. Following this a confirmatory assessment of the trainee is performed by the crew manager. As well as assessing the trainee this assessment is also serving as a check on the competence of the crew trainer in delivering the training.

This assessing of the assessor is another feature of the McDonald's approach. During one observation at a restaurant it was possible to witness a member of the crew being assessed by a crew trainer who was in turn being assessed by the franchisee (the manager of the overall operation). Just to complete the picture, a McDonald's consultant was also there conducting a performance assessment of the franchisee!

In addition to the internal workplace assessments described above, the company has instigated an external assessment process in the form of what are called 'secret shoppers'. Secret shoppers are individuals specifically recruited from outside the company to visit McDonald's outlets in the guise of normal customers and to report on such things as the cleanliness of the premises and the quality of service they received.

Overall, the aim of this emphasis on assessment in its various forms is to provide a consistently high level of performance in all employees.

Conformity with retail industry guidelines

Another part of the investigation of McDonald's assessment procedures involved a comparison between the company's training and assessment materials and the National Retail Training Package assessment guidelines. The guidelines set out requirements on such items as assessor qualifications, quality assurance and audit procedures, recording and reporting of assessment outcomes, the designing and conduct of assessments, appeals processes, recognition of current competencies, authentication and 'equal opportunity' legislation.

Assessor training and qualifications

Given the importance placed on assessment across all levels of the company structure, it is not surprising to find assessor training features in the training programs. McDonald's expects all senior members of the management team to act as workplace assessors when required, and they are required to possess at least certificate IV level training to assume this role.

McDonald's has its own set of rules for assessors which require them to:

- check information is factually correct
- ensure all areas are covered
- cover safety, harassment and other policies
- check references
- check supporting materials
- conduct workplace verification

Although somewhat different from the approach advocated by the industry guidelines, these cover most of the quality assurance, auditing, reporting and 'equal opportunity' matters specified in the guidelines.

Verification

McDonald's use a verification process to legitimate the issuing of its qualifications. The requirements for verification go well beyond those outlined in the National WRAPS Training Package.

Certificate II trainees collect a portfolio of work in their training manuals that consists of assignments and other materials generated during the training program, the most important of which are the various station observation checklists. All of these materials have been assessed and the trainee is eligible for a certificate when deemed competent in all skill and knowledge areas. However, before the certificate can be awarded, the assessment result must be signed off by the store manager or owner and finally verified by the assessment co-ordinator in the relevant State or Territory.

Design and conduct of assessments

The National WRAPS Training Package recommends two different types of assessment:

- *Integrated competency assessment:* where assessment for the Certificate II in Retail Operations is based on an integration of workplace competencies into an holistic activity. The units of competency are grouped into phases of interrelated units to facilitate the assessment process.
- *Unit assessment:* where assessment against individual units of competency is based on an integration of the performance criteria into an holistic activity for that unit.

  For certificate II the unit assessment is also seen as a means of measuring progress and may be carried out by the workplace coach/supervisor after which the integrated competency assessment becomes the formal assessment.

McDonald's approach to the design and conduct of assessment is in contrast to that recommended in the National WRAPS Training Package. Where the training package advocates a generalist and holistic approach at certificate II, McDonald's adopts a more specific one. However, beyond certificate II, assessment is more holistic.

As integrated competency assessment has not been well researched, and we know that other companies in addition to McDonald's do not always use it, the approach has been explored more thoroughly elsewhere in the report (see *Key findings*).

Recognition of current competencies/prior learning

It is possible for trainees with experience in similar companies to be 'tested out' on some modules in the program but the company-specific nature of some of the modules can make this a challenging exercise.

Authentication of work

As noted above, McDonald's assessors are required to check references, supporting materials and workplace verifications of competencies.

Quality assurance and external audit procedures

McDonald's restaurants throughout Australia are 'quality endorsed' by the relevant State or Territory training authority. Quality endorsement is only available to organisations which meet a rigorous set of criteria including a commitment to quality assurance and a process of continuous improvement.

As part of this process McDonald's undergoes an external audit of assessment procedures. These procedures are developed and managed by the training authority in the relevant State or Territory.

Equal opportunity legislation and appeals procedures

McDonald's policy on training and assessment ensures that the issues covered by equal opportunity legislation are taken very seriously. Trainees are made aware of their rights and assessors of their responsibilities. There is a training module and

assignment titled 'Responsibilities McDonald's has to its employees'. There is also an internal mediation process for dealing with appeals relating to assessment results.

Components of an assessment

A sample of the McDonald's Certificate II assessment instruments were examined to see how well they covered the essential components of competency specified in the Training Package for Assessment and Workplace Training.

These components are:

- task skills (performing at an acceptable level)
- task management skills (managing a number of different tasks that make up a job)
- contingency management skills (dealing with the unexpected)
- job/role environment skills (working with others, working as part of a team)
- transfer skills (using skills learnt in one situation in a new or different context)

McDonald's concentrated use of workplace assessment instruments makes it possible to observe most of these skills over time. However, there was limited treatment of contingency management and transfer. This, no doubt, relates directly to the responsibilities given to a member of crew. (Certificate II is essentially a crew-level qualification). Although the training and assessment program includes dealing with such things as hazards and complaining customers, the trainee is usually required to refer 'unexpected' events to his or her manager. In addition to this, contingency management skills are also a topic in the 'refreshment and reinforcement' sessions held off the job at the end of the training period.

None of the assessment instruments in the sample examined was attempting to make judgements about skill transfer. This can be related, in part, to the SOC's concentration on performances required to meet the requirements of the job at hand rather than presenting new or hypothetical contexts. Although, transfer skills were not being assessed in the sample of instruments examined there was some evidence that the importance of adapting to new and differing situations was covered in the training manuals. For example, even at certificate II level, trainees are routinely moved between different work stations and their training is designed to ensure they take the skills learnt in one station and adapt them to working in other stations.

National Pharmacies

Background

The investigations of the National Pharmacies system of assessment centred on the company's certificate II qualification, which is the minimum entry-level qualification for the retail industry. Certificate II assessment instruments related to various aspects of customer service were the focus of this part of the study.

The following information on the training programs and assessments used by National Pharmacies was supplied by the training manager who is a Certificate IV Level qualified assessor.

National Pharmacies is an organisation comprising a chain of pharmacies predominantly based in South Australia. These pharmacies are all owned and operated by National Pharmacies and all persons working in them, including pharmacists in charge, are employees of National Pharmacies. There are 31 pharmacies in South Australia, the maximum permitted by regulation. In recent years the chain has begun to extend to adjoining States so that presently there are four pharmacies in Victoria and one in NSW. In all, there are approximately 750 persons employed by the pharmacies and their parent body. As well as acting as an administration and purchasing centre for its member pharmacies, the organisation is also a registered training organisation which oversees and conducts training and assessment on behalf of its member pharmacies.

In 2000, National Pharmacies introduced a new system of training and assessment. The previous system which focussed narrowly on pharmacies, was aligned to the Quality Care Pharmacy Standards.

The new system, which is aligned to the National Wholesale, Retail and Personal Services (WRAPS) standards, was in the process of being established, pending accreditation, at the time this study was undertaken. It operates at the three levels that National Pharmacies is authorised to deliver, namely, certificates I, II and III. This study focusses on the assessments used in the new system.

A second component of the new assessment system is the Quality Care Pharmacy Program (QCPP) developed by the Pharmacy Guild of Australia and endorsed and used under licence by National Pharmacies. This program is derived from the Quality Care Pharmacy Standards mentioned above.

The assessment for each unit is usually made up of four components:

- performance of a work-related task by the candidate
- provision of documentary material by the candidate
- verification of the performance of the task, and authentication of the documentary material by the supervisor or the pharmacist in charge
- assessment of the resulting evidence by an accredited workplace assessor

The training and on-job assessment is normally conducted by the supervisor or the pharmacist in charge. However, where these personnel are not accredited assessors (which is almost always the case) the assessment evidence must be verified and signed off by an accredited workplace assessor. At present this verification is done by the National Pharmacies training manager.

The standard practice is to have the supervisor or the pharmacist in charge well enough informed and acquainted with the training and assessment procedures to perform the assessment functions themselves. Where there is any doubt the supervisor or the pharmacist in charge would be expected to contact the qualified National Pharmacies assessor for advice. The accredited assessor therefore, normally looks at the assessment documentation the supervisor or pharmacist in charge provides and makes an overall judgement. However, where necessary, he/she may visit the workplace to interview and observe the trainee, or in exceptional cases, conduct an assessment.

The last page in the candidate's principal workbook is a record page which is filled out upon completion of the assessment and a copy of it kept by the assessor as the RTO's official record. Normally the trainee submits all of the required documentation with the completed workbook. The documentation and workbook are inspected by the accredited assessor as part of the assessment and returned to the trainee as his/her record. All that is kept by the RTO is a record of the units of competence confirmed for the employee and the dates on which they were completed.

When a person comes in as a new employee they begin with the induction phase of the training program which is very comprehensively specified. The new employee's wages for the first full day of training is paid for by National Pharmacies and not the individual pharmacy. On their first day in the workplace they inspect a long list of items as an orientation to the pharmacy—including lockers, tea room, fire extinguishers, alarm buttons, and they spend an hour on the cash register. They inspect the dispensary (which is separate from the retail area), they are introduced to the concept of customer questioning, they are informed about medications at schedule 2 level and above, they look at self-care cards, they are informed about customer membership (which is peculiar to National Pharmacies) and they are showed how to fill out their pay sheets.

Within eight weeks of starting employment, the new employee is expected to have completed the first workbook. The workbook forms part of the assessment documentation. In the back is a list of the endorsed components and another of the non-endorsed components. Non-endorsed components comprise: *Introduction to the pharmacy checklist* (completed on the first day of employment); *Introduction to the pharmacy worksheet; a Quality Care Pharmacy Program* (QCPP) workbook; and a product analysis. The QCPP workbook also acts as a basic evidence guide. Employees are also given a set of 'flipper cards' with which to practise their pharmacy's required responses to various issues and contingencies. The flipper contents cover the 'team standards' which all employees are expected to follow, such as, pharmacy opening times, telephone answering procedure and rules for customer confidentiality. As well as containing standard procedures, the cards have provision for the individual pharmacy to 'customise' them by adding its own special instructions and procedures. As stated earlier, the QCPP workbook and flipper cards are used under license from the Pharmacy Guild of Australia.

As well as all of the above, new employees are given a three-day introduction to National Pharmacies which covers, among other things: merchandising, the history of National Pharmacies, marketing, membership (an important area for National Pharmacies), some OH&S, loss prevention and HR issues like awards, conditions of

service, payroll information, optical area, IT area. Then they are introduced to the retail side of the organisation and spend a morning looking at customer service followed by an afternoon with a pharmacist.

Upon successful completion of the induction phase, National Pharmacies issues the Certificate I in Retail operations. In addition to the certificate, the employee also gets a statement of attainment for the core areas in Certificate II (also referred to in WRAPS documents as Phase A) that have actually been delivered. In terms of National Pharmacies' organisation it is better for new employees to receive a certificate I to start with. This is because it is where National Pharmacies draws the line as the minimum level of competence needed to work with the organisation. By doing it that way, National Pharmacies can more easily identify employees who are motivated and have the potential to continue at certificate II level. National Pharmacies then offers employees the opportunity to take up the remainder of the certificate II. If they elect to do so they do not have to repeat the core areas (Phase A). Instead, they undertake only Phases B and C.

After completion of certificate II the employee has the option of following a retail merchandising specialist stream or being selected to follow a management stream. If they choose the former they will go on to look at an advanced merchandising and marketing course and some inventory management which gives them certificate III. If they are judged by National Pharmacies to have the potential to be a supervisor they can be put straight through to a certificate IV in Front Line Management, or they may elect to follow that path after completing Certificate III as above and serving some time on the shop floor (perhaps having been given some additional specialty retail merchandising management competencies on the way for which they can be given credit for towards certificate IV). After certificate IV, they can then be considered for selection for entry to the Diploma in Retail management.

In about eighteen months, a new WRAPS-based Retail Pharmacy Training Package providing a national system of training and assessment mutually recognised throughout the industry will be introduced. When this occurs the various pharmacy training and assessment systems presently in use, including that of National Pharmacies, will be dropped.

National Pharmacies standards and assessments

Part of the assessment review carried out with each RTO in the study involved a cross-check of the RTO standards against the assessment instruments. The purpose of this part of the review was to gain answers to the questions set out below. All findings in this study refer to assessment for Certificate II in Retail Operations.

It should be noted that there are actually two sets of standards applying to assessments conducted by National Pharmacies: one for the National Pharmacies program and another for the Quality Care Pharmacy program. As both have their origins in WRAPS they are essentially similar and no difficulties arise.

Currently, assessors do not have a separate manual or checklists. Their assessments are conducted from the documents and workbooks issued to the trainees. An assessor's guide is to be developed and used in the future.

Responses to the four questions involving the competency standards and their assessment are set out below:

- Are all the RTO's *performance criteria* covered by the assessments?

  In the National Pharmacies program the trainee is issued with a workbook which, for each unit, lists all the performance criteria stated in the corresponding WRAPS document for the unit. Alongside each of the listed criteria is a box which is ticked by the workplace assessor (normally the supervisor or pharmacist incharge) as verification that the performance of the activity has been observed and judged satisfactory. Thus all the performance criteria are assessed.

- Do the assessments cover the different situations and equipment specified in the *range of variables?*

  Although the range of variables is specified in the WRAPS standards it does not appear in the training record and assessment workbook. The tasks and other learning activities specified in the training record and assessment workbook provide reasonable cover of the variables specified in the WRAPS standards. However, the capacity of the assessments to cover an appropriate range of variables largely relies on the expertise of the assessor.

- Are the essential requirements and the underpinning knowledge, skills and attitudes as specified in the *evidence guide* covered in the assessments?

  The *critical aspects of evidence* as set out in the WRAPS standards refer to the need for *consistency* in performance of the candidate in operating point of sale equipment, applying store policies and procedures regarding point of sale transactions, handling and packaging merchandise and responsibly and accurately processing sales transaction information.

  For consistency in performance to be assessed, it is necessary for the assessor to observe the performance of the candidate on a number of occasions.

  Although the training record and assessment workbook does not specify *underpinning knowledge* and skills, they are covered to a large extent by the Quality Care Pharmacy Program, the essential elements of which are a staff member certification workbook, which sets out a series of questions and exercises relating to underpinning knowledge and skills, and a set of flipper cards which contain the information the candidate needs in order to respond correctly. The questions and activities in the workbook are assessed and formally signed off by the assessor. The knowledge and skills covered in this instrument are oriented towards the pharmacy working environment. Because of their comprehensive nature and extent of the knowledge and skills covered, the needs of a pharmacy and its employees are met admirably.

  The procedure by which the underpinning knowledge and skills are assessed in this part of the system is for the assessor firstly to evaluate the candidate's written responses to the questions contained in the workbook (the responses being obtained from the flipper cards) and secondly, to observe the performance of the candidate—usually in the form of a demonstration with the assessor acting the role of the person with whom the candidate would be dealing. Although not stated in the documents provided, National Pharmacies expects the assessor would verify the candidate's knowledge of the issues covered in the written responses by supplementary oral questions.

- Are the *key competencies* assessed separately or is the acquisition of key competencies assumed in some way?

  Inspection of the assessment materials shows that key competencies are not assessed separately. Because of the comprehensive number of performance criteria assessed, it could reasonably be assumed that assessment of most key competencies would be intrinsic to assessment of the directly work-specific competencies. For example, without competence in *Communication of ideas* the candidate could not interact competently face to face or on the telephone with customers, and without competence in Using *mathematical ideas and techniques* the candidate could not handle point of sale financial transactions competently.

Conformity with retail industry guidelines

As noted previously, the National Pharmacies training and assessment programs have been aligned with the National WRAPS Training Package. As a registered training organisation, National Pharmacies is able to customise its standards and qualifications provided they remain consistent with industry requirements.

For the purpose of this report, sections of the National Pharmacies training and assessment manuals for certificate II were cross-checked against the National Retail Training Package assessment guidelines. These guidelines require an RTO to meet a range of criteria related to such items as: assessor qualifications; quality assurance and audit procedures; recording and reporting of assessment outcomes; the design and conduct of assessments; appeals processes; recognition of current competencies; and authentication and 'equal opportunity' legislation. The findings are listed below under the relevant headings.

Assessor qualifications

Most of the assessment in certificate II is done in the workplace by supervisors or pharmacists in charge, in accordance with the training and assessment guidelines and materials provided by National Pharmacies. These personnel are subject experts rather than trained assessors, so they cannot formally sign off the assessment. Instead, they observe the candidates, inspect the material supplied by the candidates and question them to determine their competence. Having done so, they complete the paperwork and forward the completed forms, together with the accompanying documents, to the National Pharmacies assessor. This person, who is a level IV qualified assessor then inspects the evidence and if judged sufficient, signs off the candidate as competent. Where there is any doubt, or occasionally for auditing purposes, the assessor visits the candidate in the workplace for further assessment.

There are two other qualified assessors in the National Pharmacies organisation and, although they are not directly involved in the retail training area, they are available for consultation and second opinions should the retail training assessor need them.

Quality assurance and external audit procedures

At this point, there is no provision for planned internal auditing. Any such auditing that takes place is as a result of occasional visits to pharmacies by the qualified auditor to deal with training matters or assessment issues.

Apart from auditing required for RTO accreditation, there is presently no provision for external auditing. As part of the newly developed training and assessment program however, there is commitment to a review of the training program every twelve months by a panel. Although not an audit of the assessment procedures, it would seem that assessment issues could be examined.

When the new system of training and assessment was introduced, a sample of pharmacy supervisors was invited to examine the newly developed procedures and materials and report on them. The reports received were generally very favourable. One of the factors that probably contributed to this high level of acceptance was that the majority of the materials were designed with the assistance of two experienced pharmacy supervisors.

The varying levels of competence of supervisors and pharmacists in charge to gather and judge evidence were seen as an issue of concern. Because of varying levels of commitment to the task of performing these functions and to assessment in general, there was variation in consistency of assessment evidence presented to the National Pharmacies assessor for verification. This in turn placed greater load on the assessor. National Pharmacies recognises there is a need to provide training for supervisors and pharmacists in charge, and to this end is introducing a training bulletin, and plans to provide short (ten minute) training sessions on such things as what 'competent' means, what to look at, and what to look for.

Recording and reporting assessments

National Pharmacies keeps records of its trainees' assessment results on a central database. A personal file is maintained in the central office for every employee of National Pharmacies. Completed training forms are kept in these personal files. Maintenance of training records is not seen as a pharmacy function so no training records are retained by the pharmacies.

Design and conduct of assessments

The National WRAPS Training Package identifies two different types of assessment:

- *Integrated competency assessment:* based on an integration of workplace competencies into an holistic activity. To facilitate this assessment process the units of competency are grouped into phases of interrelated units.
- *Unit assessment:* assessment against individual units of competency is based on an integration of the performance criteria into an holistic activity for that unit. Unit Assessment is also seen as a means of measuring progress and may be carried out by the supervisor or pharmacist in charge after which the integrated competency assessment becomes the formal assessment.

National Pharmacies does not at this stage conduct any integrated assessment. It is the view of the training manager that the relationships between the components of integrated assessments, as they are presently structured, tend to be so weak as to

negate any advantage offered by the concept of integrated assessment. This is particularly the case for Phase A of certificate II. More credible combinations of components are needed. It was suggested that a good test of the credibility of an integrated assessment would be that all components could be covered by observation of an employee in the course of their normal duties for, say, one hour, without the need for artificially contrived work situations.

It is understood that RTO accreditation presently hinges on National Pharmacies' incorporation of integrated assessment in its assessment procedures. This issue is currently under discussion with the RTO auditor.

Appeals procedure

National Pharmacies has established processes for dealing with any grievance, concern or dispute regarding a work related issue. These matters are dealt with in accordance with the National Pharmacies personal grievance policy. If this internal grievance or appeal process does not satisfactorily resolve an issue pertaining to delivery or assessment of accredited training, the candidate may approach the South Australian Accreditation and Recognition Council (ARC) for assistance.

Recognition of current competencies/prior learning

National Pharmacies values the knowledge and experience individuals bring with them when joining the organisation. To this end, there is a National Pharmacies policy on recognition of prior learning and recognition of current competency. Assessments for RPL and RCC are conducted by the National Pharmacies qualified assessor.

As an interesting sideline to this issue, it was commented that training and assessment received under the National Pharmacies system might not always receive recognition by other pharmacies which used an alternative system of training due to the competitive nature of the training market. Of course, completed WRAPS qualifications would still have to be recognised as such, but they may not necessarily be accepted as sufficient for employment in a pharmacy. As stated earlier in this report, in about eighteen months, a new Retail Pharmacy Training Package, presently under development, will be introduced. With its implementation will come a national system of training and assessment which will be mutually recognised throughout the industry, so this issue should then no longer arise.

Authentication of work

Assessment in the National Pharmacies program includes making judgements on some work done outside of normal work time and which, therefore, is unsupervised. As unsupervised work does not represent a very significant proportion of work assessed, the need for authentication of work presented for assessment is not great, but the need does exist. It is expected that supervisors would already be informally authenticating unsupervised work, however, National Pharmacies recognises this as an issue and will be looking at the possibility of developing policy on authentication in due course.

Equal opportunity legislation

Equal opportunity in regard to assessment is covered by the Equal Opportunity component of the National Pharmacies Human Resources policy.

Components of an assessment

A sample of the National Pharmacies Certificate II assessment instruments were examined to see how well they covered the essential components of competency specified in the Training Package for Assessment and Workplace Training.

These components are:

- task skills (performing at an acceptable level)
- task management skills (concurrently managing a number of different tasks that make up a job)
- contingency management skills (dealing with the unexpected)
- job/role environment skills (working with others, working as part of a team)
- transfer of skills (using skills learnt in one situation in a new or different context)

The National Pharmacies' emphasis on workplace assessment makes it possible to observe the trainee's level of performance in the first four components over time. The assessor checklists give an adequate coverage of skills required to perform tasks, manage tasks, deal with the unexpected and work in a team. However, one component which did not appear to be explicitly covered was transfer of skills.

Within a pharmacy any skills transfer that does occur is likely to be of a minor nature and rather subtle, which limits the opportunity to assess it. However, opportunities do exist. Examples of skills transfer which might be observed and assessed are: applying product display skills learnt in one store section to tasks undertaken in another section, applying product knowledge in different contexts, or applying skills learned in dealing with one customer to dealings with other customers.

DJ Foreman & Associates

Background

David J Foreman & Associates (DJFA) is a consultancy based in Adelaide and operating across Australia on its own account and in conjunction with a network of private consultancies located in all major capital cities.

DJFA is a registered training provider for the qualifications of:

- Workplace Trainer (Category 1)
- Workplace Assessment training
- Certificate IV in Workplace Assessment
- Certificate IV in Workplace Training and Assessment

In addition, DJFA provides other training in the areas of management, personal development and computing.

Information on DJFA was provided by its managing director who has a Diploma in Training and Assessment Systems.

The company delivers most of its assessor training within the private sector. As well as the training of assessors, DJFA has been a leading contributor in national projects involving the development of assessment procedures in a number of industries and sectors including:

- National Occupational Health and Safety Commission
- forest industry
- electrical industry (Restricted Electrical Licensing Committee)
- natural resource management

The investigations of David J Foreman and Associates were centred on the company's assessor training and, in particular, the procedures used to assess the competence of assessors.

David J Foreman & Associates Standards and assessments

In order to become a registered training organisation DJFA was required to demonstrate to the Accreditation and Registration Council (ARC) in South Australia that it met the national principles, standards and protocols that make up the Australian Recognition Framework.

This involved:

- *Meeting the core standards required to follow national principles for registration and mutual recognition namely: following all relevant law; having a commitment to access and equity; demonstrating a focus on quality; agree to external monitoring and audit; having sound financial, administrative and client/learner records management systems; and engaging in ethical marketing and advertising.*
- *Possessing the training products and services to meet the requirements of the Training Packages for Assessment and Workplace Training; being able to identify the learning needs of diverse types of clients; and being able to conduct assessments and issue qualifications and statements of attainment.*

- *Agreeing to conform with protocols covering compliance audits and strategic evaluations, ethical marketing and registration fees.*

Registration for the SA Accreditation and Registration Council gives DJFA national recognition to train and assess the courses set out in the Training Package for Assessment and Workplace Training. Since achieving RTO registration in 1994 the company has delivered assessment courses in the ACT, Western Australia, Victoria and South Australia.

The review examined the relationship between the industry standards set down in the training package and DJFA's assessment procedures. The company provides trainees with copies of the relevant course units including *elements of competence, performance criteria, range of variables* and *evidence guides*. As well as making the training process transparent in this way, the trainees are also encouraged to identify areas within the program where they might be able to gain recognition for prior learning.

Because the company conducts training for assessors and workplace trainers in a range of industries and across a variety of enterprises, it has found it necessary to tailor its program to meet individual client needs. Consequently, there is no one approach to either the training or its assessment.

It was, however, possible to examine the approach used in assessing the three benchmark units needed by assessors ('Plan assessment', 'Conduct assessment' and 'Review assessment').

Integrated assessment and standards

DJFA adopts an integrated (holistic) approach to assessing these units by gathering and judging evidence that may relate to all three units in a single assessment event. Assessments are conducted on several occasions during a program. This interdependent approach to assessment is endorsed in the training package.

The challenges presented by the integrated approach to assessment are discussed elsewhere in the report (see *Key findings*) but, in the case of DJFA, the company's checklist could be compared against the requirements listed in the evidence guides.

Allowing for some differences in the wording on the DJFA checklist, all of the evidence requirements listed in the training package seemed to be covered.

For its part, DJFA includes *Assessment of key competencies and Appeal procedures* that, although referred to in the training package, are not specifically listed as evidence requirements.

Assessment methods

The assessments used by DJFA include:

- written exercises during training
- short, written paper near the end of the formal training
- assessment practicals during training
- an assessment project (preferably in the workplace)

In cases where this battery of assessments does not allow the trainees to demonstrate all of the competencies with the required consistency, they are required to complete additional assignments in the workplace.

Overall, the DJFA assessments have a strong workplace focus but not to the detriment of underpinning knowledge requirements. In particular, the project work is designed to provide the opportunity to gather evidence on virtually the whole gamut of performances required of a competent assessor including:

- specifying the competencies to be assessed
- designing the assessment methods
- trialing the methods
- using RPL/RCC assessment as appropriate
- negotiating the process with the person to be assessed
- conducting the assessment
- recording, reviewing and providing feedback to the person assessed

The reviewer noted that although the DJFA assessment checklist did not include an item on 'developing and validating assessment tools', these competencies were covered in the assessment project.

Conforming with training package and training authority guidelines

The Training Package for Assessment and Workplace Training sets out a number of guidelines for the conduct of assessments. However, because registration is the responsibility of the individual States and Territories, it is the South Australian guidelines that sometimes take precedence.

Code of practice

The training package makes reference to an international code of practice for assessors developed by the American National Council for Measurement in Education (NCME). However, DJFA has adopted the standard code of practice specified by the SA Accreditation on Registration Council. (This is the Council's requirement for achieving RTO status.) The code lists the requirements to be met by an RTO and requires a substantial amount of information to be provided to trainees before the trainees enter into a contract of training. A list of these requirements is provided as attachment A (p.76).

National Assessment Principles

In a similar treatment to that given to the code of practice for RTOs the SA Accreditation and Registration Council have also provided an interpretation of Training Package's National Assessment Principles. The principles which are provided as attachment B (pp.77–78) spell out the requirements applying to an RTO in South Australia.

DJFA trainers and assessors work to the principles laid down in attachment B and a copy of these is provided in the company's trainers and assessors manual.

Recognition of current competencies/recognition of prior learning

The training package guidelines do little more than acknowledge that the use of RCC/RPL procedures involves a greater reliance on indirect or supplementary forms of evidence (for example, portfolios of work, formal qualifications and references). DJFA, on the other hand, see RCC/RPL procedures as one of the more important methods of gathering assessment evidence. The company's manual for its trainers and assessors puts substantial stress on RCC/RPL procedures and provides additional advice on access, trainers and resource issues as well as explaining what makes RCC/RPL a different type of assessment.

Appeals procedure

Depending on the form of appeal, the company delegates either the managing director or another director to hear the appeal. Procedures are specified in the company manual and conform to the requirements of the ARC. In the six years of operation there has never been a need to use these procedures.

Authentication of work

Authentication is an important issue for DJFA because of the substantial use the company makes of RCC/RPL procedures. The checks carried out depend on the nature of skills being assessed, but normally involve oral questioning of the trainee with further questioning of peers and interviews with relevant individuals, including previous employers, when dealing with evidence of an historical nature.

Quality assurance and external audit procedures

DJFA has not chosen to apply to become a quality endorsed organisation. Whether the self-managing status which comes with quality endorsement would have substantial benefits to the business operations of the company is a matter for consideration in the future. Any move towards endorsement would need to consider how this might impact on the relationships it currently enjoys with consultancies in the other States.

Nevertheless, although quality is dealt with in the National Assessment Principles and DJFA's manual thoroughly documents the system and procedures that go with quality assurance, there is no reference to the company's policy on quality. The manual would be enhanced if the point were made that quality is not simply about keeping the stakeholders and the clients satisfied through good record-keeping, accountability and consistency of performance. More than anything else, it is about continuous improvement through a cyclical process of planning, doing, checking and improving.

External audits to ensure compliance with the requirements of registration are conducted by the Accreditation and Recognition Council. Since its registration as an RTO in 1994, DJFA has been formally audited on two occasions, the most recent being early in the year 2000.

Components of an assessment

An example of DJFA's assessment checklist for the three units—Plan, Conduct and Review assessment was examined to see how well it covered the essential components of competency specified in the Training Package for Assessment and Workplace Training.

The training package provides examples of how evidence can be obtained for each component as shown below.

Component	**Example of possible evidence**
Task skills *(performing at an acceptable level of skill)*	Performance of assessment and training competencies as observed and interpreted against the performance criteria
Task management skills *(managing a number of different tasks within the job)*	Demonstration of how a number of inter-related assessment and training competencies are managed
Contingency management skills *(responding and reacting appropriately to unexpected problems, changes in routine and breakdown)*	Demonstration of appropriate personal management and other adjustments to training and/or assessment procedures when the group, materials and/or candidates are non-routine or not as expected
Job/role environment skills *(fulfilling the responsibilities and expectations of the workplace)*	Demonstration of how assessment and training competencies are incorporated into workplace responsibilities and realities
Transfer skills *(transferring skills and knowledge to new situations and contexts)*	Demonstration of assessment and training competencies in unfamiliar contexts that have not been used in the learning or specific workplace application (NAWTB 1999, p.15)

However, with the exception of task skills, none of the other examples can be found in the evidence guides for the three units. This creates a serious problem for users of the package, especially as the approach outlined is meant to show how integrated assessments can be designed.

The DJFA course assessments, in particular the project work, with its workplace emphasis gave a reasonably good coverage of the five components. However, just how good a coverage will depend on the project chosen. DJFA gives the trainees some guidance in this regard but it could be strengthened by specifying what must be done to include each component in the project or practical work.

Attachment A

Information to be provided to trainees by the registered training organisation before trainees enter into a contract with them:

- copy of the code of practice
- the accreditation status of the course
- entry requirements
- arrangements for the recognition of prior learning
- the commencement dates and duration of courses
- the time commitment involved in undertaking the training offered
- the qualifications/certification to be issued on completion or partial completion of the course of study
- requirements to achieve the qualification
- how the course articulates with other training
- expected employment outcomes
- policies on assessment, grading, resubmission of work etc
- detailed costs of training
- the conditions under which trainees will be eligible to receive a refund of fees
- arrangements for the protection of students' funds
- internal and external grievance/appeal processes
- students' rights and responsibilities
- RTO's rights and responsibilities
- withdrawal arrangements
- conditions under which tuition may be terminated
- trainee support services

Attachment B

National assessment principles

This section outlines the set of National Assessment Principles which have been endorsed, in principle, by the State and Commonwealth Ministers responsible for vocational training. The principles provide the basis for the ongoing development of assessment systems and practices in vocational education and training. Following each principle is a brief explanation of its intent and application in South Australia.

Principle 1—Endorsed industry/enterprise standards form the basis of qualifications in the vocational education and training sectors, where they exist.

Qualifications are approved through the endorsement of training packages by the National Training Framework Committee (NTFC). Training packages comprise endorsed industry standards, assessment guidelines and qualifications defined simply by the specification of which units of competence make up each qualification. Course accreditation only applies where no relevant qualification in endorsed training packages exist.

Principle 2—Endorsed industry/enterprise standards are the benchmarks for assessment, where they exist.

An RTO should assess trainees directly against the units of competence which make up a qualification

Where no industry endorsed standards exist, the outcomes specified in the curriculum for an accredited course will form the benchmark for assessment.

Principle 3—Assessment conducted for the purpose of national recognition should lead to a part or full qualification under the Australian Qualifications Framework.

Trainees who are assessed and meet the requirements for one or more units of competence/modules will receive a statement of attainment from the RTO indicating the units/modules attained. This statement of attainment will be recognised by all other RTOs.

Principle 4—Assessment should be undertaken by, or auspiced through, a Registered Training Organisation

Only assessments authorised and recorded by an RTO will be recognised in the VET system. Other parties may contribute evidence to the assessment decision made by the RTO. Conversely, organisations delivering training who do not wish to be registered may form partnerships with RTOs who will be responsible for overseeing assessments and issuing statements of attainment or qualifications. An example of this is the VET in SACE Arrangements (VISA).

Principle 5—Assessment for national recognition purposes shall be conducted within a quality assurance framework

An RTO must ensure assessment decisions are made by people who are qualified to conduct the assessments. Acceptable arrangements for the conduct of assessments will be described in the assessment guidelines in the training package or in the curriculum document.

An RTO will be reviewed by the ARC in conjunction with industry representatives to ensure assessment processes are appropriate.

The national industry training advisory board (ITAB) network will be responsible for reviewing the industry endorsed standards and assessment guidelines to ensure these are facilitating good assessment decisions.

Principle 6—Responsibility for assessment resides with the body that issues the qualification under the Australian Qualification Framework.

The ARC will authorise an RTO to issue qualifications. An RTO will be responsible for ensuring the full requirements of the qualification have been met. This may include recognising statements of attainment which contribute towards the qualification issued by other RTOs.

An RTO may form a partnership with another organisation/enterprise that contributes to the assessment process to promote pathways to recognition for clients and to assure quality outcomes.

Principle 7—Assessment processes shall be valid, reliable, flexible and fair.

Features of quality assessment systems include, but are not limited to:

- competent assessors
- authentic workplace conditions (this does not exclude the use of simulated workplace conditions unless actual workplace conditions are specified in the evidence guide of the unit of competence)
- integration of skills and knowledge
- integration of task management and contingency management skills
- adequate information provided to trainees prior to and on completion of assessment
- flexibility in the methods of assessments to accommodate individual needs without compromising standards

Logistics Training Unit—Douglas Mawson Institute

Background

The Logistics Training Unit is a registered training organisation (RTO) of the Douglas Mawson Institute of TAFE in South Australia providing training to the transport and distribution sectors of industry. The following information on the training programs and assessments used by the Logistics Training Unit was supplied by the education manager who is a certificate IV level qualified assessor.

The Logistics Training Unit presently provides training for warehousing qualifications in South Australia, the Northern Territory, New South Wales and Queensland. The training is provided both in training centres and workplaces, with a significant proportion of the training being delivered through distance education. Assessments are normally in the workplace and on the job, with Logistics Training Unit staff frequently travelling interstate for this purpose. The Logistics Training Unit offers training and assessment in the fields of road and rail transport as well as warehousing.

The Warehousing training programs and assessments which the Logistics Training Unit provides are aligned to the Transport and Distribution Training (Warehousing) package produced by the Transport and Distribution Training Australia with funding support from the Australian National Training Authority (ANTA). They extend from certificate I level through to diploma. The Certificate II in Warehousing is the focus of this report with special attention being paid to Unit D12 97A: *Operate specialised load handling equipment*.

The primary documents used by the trainees are a learner's guide and a training and information record (generally referred to as the logbook) for each level. The introductory section of the learner's guide contains advice to trainees on how to plan their training together with explanations of the nature and content of what they will learn, the concept of competency standards, and the way in which they will be assessed. Following this, there are sections devoted to underpinning knowledge, suggested learning activities, short-answer quizzes and checklists, all covering a number of the range of variables items and performance criteria. There are also details of some assessment tasks the trainee will be asked to perform. In the logbook, the methods of assessment are clearly stated.

In conducting an assessment, the assessors work from the learner's guide and the training and information record as issued to the trainee. No separate guide for assessors is provided.

Industry standards and Logistics Training Unit assessments

Part of the assessment review carried out with each registered training organisation in the study involved a cross-check of the assessment instruments against the relevant industry standards. In the case of warehousing, the standards used are contained in the Transport and Distribution Training Package.

The purpose of this part of the review was to gain answers to the questions set out below. All findings refer to assessment for the Certificate II in Warehousing.

- Are all the RTO's *performance* criteria covered by the assessments?

  The Performance Criteria are specified in detail in the Transport and Distribution Training standards. However, they are only partially specified in the learner's guide (also known as the trainee's workbook). Typical tasks in which they are applied are suggested in the trainee's logbook together with the elements of competence, but without guidance from a trainer, the trainee could not be expected to be aware of precisely what was being looked for in the assessment. The trainee's learner's guide and logbook, which the assessor works from in conducting the assessment, are normally the principal sources of information for assessors regarding the performance criteria. In some instances, performance criteria may also be specified in additional checklists provided by the trainee's on-job supervisor. The tasks and other activities specified in the learner's guide should enable the performance criteria to be covered.

- Do the assessments cover the different situations and equipment specified in the range of *variables?*

  As was the case for performance criteria, the range of variables were clearly stated in the Transport and Distribution Training standards. The range of variables are indirectly, but reasonably well covered from a trainee's perspective by the lists of equipment and the range of activities suggested in the learner's guide. Information provided to assessors concerning the range of variables is essentially that which is contained in the learner's guide and logbook.

- Are the essential requirements and the underpinning knowledge, skills and attitudes as specified in the *evidence guide* covered in the assessments?

  The evidence requirements specified in the Transport and Distribution Training standards evidence guide are not directly specified in the learner's guide or logbook. Although the activities, instructions and explanations contained in the learner's guide should enable all assessment evidence requirements to be met, the assessor would need to refer to the standards to obtain a clear indication of what was required.

- Are the *key competencies* assessed separately or is the acquisition of key competencies assumed in some way?

  Because of the holistic nature of the assessments, the key competencies are integrated into the units of competency and not assessed separately. However they are specified separately in the competency standards together with the levels expected.

Comment

The Logistics Training Unit believes one of the strong features of this training program and its assessments is its transparency, which is to say, nothing is hidden from the trainee. Inspection of the training materials generally supports this view, however, as pointed out above, the performance criteria would not be evident to the trainee. Ideally, everything the trainee is expected to know and be able to do should be covered in the learner's guide. If transparency is to be complete therefore, the performance criteria should also be covered.

Conformity with warehousing industry guidelines

As noted earlier, the Logistics Training Unit training and assessment programs have been aligned with the Warehousing Training Package. As a registered training organisation, the Logistics Training Unit is able to customise its standards and qualifications, provided they remain consistent with industry requirements.

For the purpose of this report, sections of the Logistics Training Unit training and assessment manuals for certificate II were cross-checked against the Warehousing Training Package assessment guidelines. These guidelines require an RTO to meet a range of criteria related to such things as: assessor qualifications; quality assurance and audit procedures; recording and reporting of assessment outcomes; the design and conduct of assessments; appeals processes; recognition of current competencies; authentication; and 'equal opportunity' legislation. The findings are listed below under the relevant headings.

Assessor qualifications

All personnel conducting assessment in the Logistics Training Unit Warehousing training program are either qualified level IV assessors or are in training for a level IV assessor qualification. There are currently twelve level IV qualified assessors in the logistics training unit plus two in training for level IV.

Assessments are also undertaken by subject experts in collaboration with a qualified assessor.

Quality assurance and external audit procedures

The Logistics Training Unit undergoes an audit every six months by the Accreditation and Registration Council (ARC) of SA. This is an audit of compliance with administration procedures and records rather than of the assessors and assessment processes.

External audits of assessments take the form of quality assurance feedback from employers. This feedback is necessary to satisfy ANTA requirements. When an assessor visits an employer to check progress of trainees and administer assessments a 'site visit and progress report' is prepared. On it, the employer is invited to report on the trainee's progress in terms of: skill development, attitude, and application. They are also invited to report on their satisfaction with the service provided by the Logistics Training Unit staff in terms of: mentor skills, industry knowledge, and timing and duration of visits. While there is no specific mention of assessments, it is expected that feedback would also cover these.

From an assessment auditing standpoint, it would be desirable to have assessments specifically covered in the 'site visit and progress report'.

As nearly all the assessors are either level IV qualified or in training for level IV and work under the leadership and guidance of personnel who have a background in workplace training, there are no serious concerns about the quality of the assessors or assessment procedures. In addition, although there is no formal internal auditing of assessors, often two assessors will go to a centre or workplace together to conduct assessment (as many as 30 individual assessments is not uncommon) and

they will actively observe each other during the assessments—in effect, validating each other. This established practice helps produce highly skilled assessors.

Recording and reporting assessments

Assessment results are recorded and reported under the SA Department of Education, Training and Employment (DETE) Student Management System (SMS) and under the National VET Management Information System as well. One of the advantages in having access to the SMS is that it enables reliable, long-term records to be generated and maintained.

On visiting trainees, assessors generate reports which can be for a group or an individual trainee. In normal circumstances these reports merely document the progress of the trainees in terms of readiness for assessment and, where a competency has been achieved, the date on which the trainee was assessed as competent. Where there are special issues these are recorded as additional notes on the forms. These issues can be things like a slow rate of progress of a trainee or group of trainees, need for additional training or special coaching, any action needed by the employer (for example, provide experience on particular machines), and any special issues to be taken into consideration regarding future assessments.

A new form is completed by the assessor on each visit and a copy is incorporated in the Logistics Training Unit records. An additional copy may also be given to the employer, particularly when it details any suggested action required of the employer.

Design and conduct of assessments

In some training packages two different types of assessment are identified. They are:

- *Integrated competency assessment:* based on an integration of workplace competencies into an holistic activity. To facilitate this assessment process the units of competency are grouped into phases of interrelated units.
- *Unit assessment:* assessment against individual units of competency is based on an integration of the performance criteria into an holistic activity for that unit.

Unit Assessment is seen by some as a means of measuring progress and may be carried out by a supervisor after which the integrated competency assessment becomes the formal assessment.

Integrated competency assessment is not written into the Transport and Distribution Training Package and is not used as such in assessments conducted by the Logistics Training Unit. However the holistic nature of the assessments conducted makes them similar to integrated assessment.

Appeals procedure

There is an appeals process; however, it is not specified in the training package.

The process used is the standard SA TAFE appeals process which, in effect, consists of three stages:

- the trainee endeavours to resolve the issue through discussion with the assessor
- the trainee goes to the educational manager in charge of the assessment (possibly resulting in re-assessment by another assessor)
- the matter goes to a panel for discussion and resolution

Recognition of current competencies/recognition of prior learning

RCC is applied at levels I to V; however, at levels IV and V it becomes a combined RPL and RCC assessment because of the larger theory component at these levels. In this context, RPL is seen to be limited essentially to credit transfer, whereas RCC is recognition of prior experience as well as credit transfer.

At level VI the type of recognition pursued depends on the reason for requesting it; for example, if the trainee wants to go on to the University of South Australia it has to be RPL (essentially a completed associate diploma) the university will not accept RCC. The Northern Territory University, on the other hand, will recognise RCC.

Authentication of work

Authentication procedures are primarily used in assessment for RCC. The stringency of authentication is, to some extent, dependent on the training level. At levels I to III, if there is any cause to doubt the authenticity of a piece of work, the assessor will ask questions of the trainee or ask the trainee to perform the task. At levels IV to VI authentication is nearly always required. The standard practice in these cases consists of three stages:

- the applicant is required to submit a self assessment. If this is not sufficient,
- a third party verification is called for, and if not sufficient,
- the Logistics Training Unit undertakes an assessment of its own

There was little reference in the training and assessment materials used by students and assessors to the need for authentication of trainees' work as part of normal workplace assessment. For example, authentication which might be done by a supervisor to confirm that work completed by the trainee for inclusion in the learner's guide was actually done by the trainee. There was an expectation that the assessor would question the trainee, where necessary, to establish authenticity. However, there was no direct reference to the need for authentication of trainees' work or the procedures used for authentication either in the learner's guide or the logbook which are the assessment documents used both by the trainee and the assessor.

Equal opportunity legislation

Equal opportunity is not an issue in Logistics Training Unit assessments in warehousing. All trainees have to perform the same tasks irrespective of age, gender or any other factor.

There is also a safeguard in that assessment must satisfy the broad requirements of the SA TAFE equal opportunity policy with which the Logistics Training Unit is obliged to comply.

Components of an assessment

A sample of the Logistics Training Unit assessment instruments for Certificate II in Warehousing was examined to see how well they covered the essential components of competency specified in the Training Package for Assessment and Workplace Training.

These components are:

- task skills (performing at an acceptable level)
- task management skills (managing a number of different tasks that make up a job)
- contingency management skills (dealing with the unexpected)
- job/role environment skills (working with others, working as part of a team)
- transfer of skills (using skills learnt in one situation in a new or different context)

With its strong emphasis on assessment on the job, the Logistics Training Unit system of assessment should be well suited to assessment of all five of the components listed above. However, the great degree of flexibility allowed in the choice of assessment tasks and the strongly holistic approach to defining and assessing them places a great deal of responsibility on the assessor to ensure that all five components are actually addressed.

As mentioned earlier, assessors do not have a separate assessor's guide or checklist. Coverage of the above components in an assessment is therefore largely dependent on the experience and expertise of the assessor.

Other issues

Grading

Because all assessment is competency based, no grading is used. However, it was recognised that some aspects of trainee performance could be used to grade them. For example the time taken to complete a unit could indicate a trainee's aptitude for that type of work. If a person gets through it faster, then it might be assumed they would have a higher level of capability in that area. More generally, the time taken to complete a whole program of units could be an indicator of overall level of competence in the work. Some students, for example, have completed a certificate III in six to eight months, when the normal time taken is eighteen months or more. At the other end of the scale, the number of attempts to achieve competency could also serve as a means of grading. Of course enthusiasm and degree of application also play a part in any outcome.

Scheduling of assessment

Logistics Training Unit personnel have also observed that the time of day or the time within a shift at which an assessment is conducted can also have a bearing on the result. They have found trainees perform better at the beginning than at the end of a shift, and that 2:00 pm in the afternoon is a particularly bad time to conduct an assessment, there being a noticeable drop in performance even from 1:00 pm to 2:00 pm.

Cost of assessment

The Logistics Training Unit has also developed some unofficial estimates of the cost of assessment. For an independent assessment of a unit of competency for one person, not forming part of a training program, the cost is approximately $150. If included as part of a traineeship, the cost of an assessment is around $20, based on the fact that it takes about half an hour to complete an assessment at a rate of approximately $40 per hour. The charge for assessment of a trainee is less than that for independent assessment outside a training program because much of the trainee's assessment evidence is checked by the supervisor rather than by the Logistics Unit assessor.

The Logistics Training Unit has also costed RCC. For levels I, II and III it is $10 per unit of competency, or $70 for a full qualification. Level IV, which has more depth, is $25 per unit of competency and at diploma and advanced diploma levels, it is $50. These costs were based on matching the amount of time spent on the activity with an ASL 1 (Lecturer Grade 1) salary. The actual fees charged by the Logistics Unit for RCC are the same as the typical costs specified above. They are set charges and are not adjusted for variations in time spent on the assessment.

Quality of evidence is an important factor in determining the true cost of RCC assessment. By way of example, a recent RCC assessment at level IV took only half an hour because the candidate was in the army and the armed services records were comprehensive. By comparison, if good records are not available, a similar assessment could take as much as six hours. An interesting sidelight to the army candidate was that the RCC assessment was conducted for a person in Darwin by an assessor in Adelaide using electronic communication—an indication of change taking place in the field of assessment.

Retail Training—School of Business Studies

Regency Institute of TAFE—Elizabeth Campus

Background

The Retail Studies program is offered by the Business Services Program of the Regency Institute of TAFE in its capacity as a registered training organisation (RTO). The Business Services Program also offers programs in other business-related areas such as administration studies, accounting, management, small business management and information technology, as well as auspicing the VET-in-Schools program for a number of local schools. The latter program enables school students to qualify for Certificate I in Retail Operations and in some cases the Certificate II in Retail Operations while still at school.

Training and assessment in the retail studies area ranges from Certificates I to III in Retail Operations through to the Certificate IV in Retail Management and the Diploma in Retail Management. Certificate I is not offered at TAFE as a stand-alone qualification as such, but forms Phase A of the certificate II program. Students entering the certificate II course with the certificate I are given credit for Phase A of certificate II.

Retail organisations accommodated by the Business Services Program range from one-person businesses with one employee to large organisations. Much of the work is currently with small-to-medium-size businesses. In the retail industry, many of the larger organisations like Woolworths and Coles are registered training organisations in their own right and do most of their own training and assessing.

The majority of training and assessment in retail studies is offered in the comprehensive training facility at the Elizabeth Campus, which will be described in greater detail below. An increasing amount of assessment and some training is conducted in the workplace. All assessment is against the competency standards found in the National Wholesale Retail and Personnel Services Training Package. In the case of clients involved in on-the-job traineeships the RTO's responsibility is to verify that the trainee has been assessed against the specified competency standards. Regency workplace assessors regularly visit trainees on the job to verify that assessment has been properly conducted. This program forms part of the New Apprenticeships system.

Regency Institute also has a second client group which require training and assessment to be conducted within the institute. This group do not have employment, that is they are not under contracts of training. For these students the institute has a retail training centre, operating under the name of 'Taffy's Department Store', set up as a retail store with goods, cash registers, point of sale equipment, displays, departments, telephone service, reception, and other retailing equipment which meet industry standards. Within the training centre, students have to set up displays, promote and sell products, conduct sales, handle complaints, receive and fill telephone orders, buy in stock and deal with other departments.

There is also a model office where administration studies students can learn reception skills, filing, shredding, photocopying and other office practices. Retail

students also have access to this model office. As of Semester 3 in 2000, the institute began to operate a 'practice firm' which will work in conjunction with the retail training centre and operate jointly under the name of 'Taffy's Department Store'. Although initially intended for administration studies students, it will also be used by students from the retail training program. Practice firms are 'virtual firms' utilisng during training, all normal organisation practices including paperwork and using Australian Network of Practice Firms cheques. Actual goods are not traded. The Practice Firm has a typical range of departments such as finance, reception and purchasing, marketing, and human resources.

The Practice Firm Network operates nationally and internationally. Virtual trading takes place between all practice firms. The mentor for the Regency Institute's Practice Firm is Harris Scarfe, a major SA department'store chain which serves as a model and assists and advises as well as donates equipment for the institute's retail training centre. It is expected that the Regency Institute's practice firm will 'specialise' in corporate wardrobes and the like. Students rotate within the various 'departments' during their training.

Students who have not had any retail experience at all are encouraged to undertake a one day per week industry placement as part of their training. Some who have only had retail experience in a narrow area may also choose to undertake work experience in other areas. A number of the students who participate in this program obtain part-time, casual, or full-time work.

As the WRAPS packages are primarily designed for on-job training and assessment, the institute has designed assessments which match job requirements. The instructors the institute uses in the retail program are part-time trainers recently recruited from industry.

In certificate II, which is the focus of this report, the institute has 20 to 25 students each semester. Each of these students undertakes 10 *essential* units and 3 *elective* units, covering 13 competencies in all. For each unit the students follow a learner's guide, complete various field trips, and undergo a series of observations and verifications. The 13 competencies are assessed by means of unit assessments and three integrated competency assessments as specified in the WRAPS package: the assessment for Phase A covers five competencies, Phase B, seven and Phase C the remaining three.

Industry standards and Business Services Program assessments

Part of the assessment review carried out with each registered training organisation in the study involved a cross-check of the assessment instruments against the relevant industry standards. The standards against which the assessments were checked in this instance were those specified in the WRAPS National Retail Training Package.

The purpose of this part of the review was to gain answers to the questions set out below. All findings refer to assessment for the Certificate II in Retail Operations.

- Are all the RTO's *performance criteria* covered by the assessments?

  Although the Performance Criteria are not specified in the assessment documents, they are covered by the listed tasks the trainee must undertake

(which are specified in some detail). All students and trainees have a training record book that specifies the units of competency, the elements of competency and the performance criteria for each competency. All students and trainees therefore have a clear idea of the performance criteria against which they are being assessed.

- Do the assessments cover the different situations and equipment specified in the *range of variables?*

  The tasks and other learning activities undertaken by the trainees provide the necessary coverage of the variables specified in the WRAPS range of variables. The training program is designed to be flexible and to meet the needs of the trainees in terms of where and when training and assessment occurs and the type of assessment undertaken.

- Are the essential requirements and the underpinning knowledge, skills and attitudes as specified in the *evidence guide* covered in the assessments?

  The assessment instruments are designed to cover all the Evidence requirements of the Retail Training Package. The tasks and other learning activities undertaken by the trainees cover the requirements of the standards regarding underpinning knowledge, skills and attitudes.

- Are the key competencies assessed separately or is the acquisition of key competencies assumed in some way?

  Because of the holistic nature of the assessments, assessment of key competencies is not undertaken separately but is an inherent part of the assessments undertaken by the trainees.

Conformity with retailing industry guidelines

As noted earlier, the retail training program assessments conducted as part of the Business Services Program have been aligned with the WRAPS Training Package.

For the purpose of this report, sections of the Business Services Program training and assessment manuals for certificate II were cross-checked against the WRAPS Training Package assessment guidelines. These guidelines require an RTO to meet a range of criteria related to such items as: assessor qualifications; quality assurance and audit procedures; recording and reporting of assessment outcomes; the design and conduct of assessments; appeals processes; recognition of current competencies; authentication; and 'equal opportunity' legislation. The findings are listed below under the relevant headings.

Assessor qualifications

Assessors are qualified Level IV Workplace Assessors, meeting the requirements of the state Accreditation and Registration Council (ARC). All have either worked in or are currently employed in the retail industry.

Quality assurance and external audit procedures

As the institute has a 'quality endorsed' status from the State training authority it is required to have quality standards in place with respect to all aspects of the training product. Independent auditors may review such products at any time for ongoing

recognition. The co-ordinator is also a member of the TAFE Retail Studies Quality Assurance Group which has members representing each institute in which Retail Studies are conducted. Members of this group discuss and compare what they are doing in relation to assessment. Trainers/assessors also have regular meetings to discuss standards. All assessment instruments are checked by at least three people.

To maximise flexibility of the training program for the trainees, up to three attempts to rectify minor problems are allowed before a trainee is required to undertake further practice or study before being assessed via a new assessment instrument and a new situation. In retail studies, a successful strategy has been to advise trainees which areas of training need attention to enable them to concentrate their effort on improving in those areas. It has also been found that trainees who realise they are unable to cope withdraw from the program of their own accord.

Lecturer/assessors do everything possible to ensure that students/trainees are able to complete their assessments within the semester, often to the extent of assessing students outside their paid time. In addition, students/trainees are offered the opportunity to attend additional classes at no extra expense.

Recording assessments and grading

Assessment results are recorded and reported under the departmental Student Management System (SMS) and under AVETMISS as well. One of the advantages in having access to the SMS is that it enables reliable, long-term records to be generated and maintained.

Grading is undertaken only at certificate IV and diploma levels. It is based on additional criteria such as demonstration of wider reading, ability to work unsupervised, submissions which require a minimum of amendment, demonstration of original and independent application of theory and practice, and ability to interpret and express complex ideas clearly. Students may be given a distinction, credit or pass but not a numerical grade such as a percentage.

Grades are usually provided in response to stakeholder's requests (such as to satisfy university selection criteria).

Further discussion of grading may be found in key findings in the main report.

Design and conduct of assessments

In the WRAPS Retail Training Package two different types of assessment are identified. They are:

- *Integrated competency assessment (ICA):* based on an integration of workplace competencies into a holistic activity. To facilitate this assessment process the units of competency are grouped into phases of interrelated units.
- *Unit assessment:* assessment against individual units of competency is based on an integration of the performance criteria into an holistic activity for that unit. Unit assessment is seen by some as a means of measuring progress and may be carried out by a supervisor after which the integrated competency assessment becomes the formal assessment.

Off-job trainees complete assessments of individual units and integrated competency assessments. On-job trainees complete ICAs and are also assessed by

their workplace supervisor or coach. The on-job assessments are verified by a qualified assessor from the Regency Institute.

Assessments are largely competency-based with approximately 90% of assessment being undertaken on a one-to-one basis. Trainees are therefore closely observed putting skills and understanding into practice rather than reiterating what they have read or been told. Trainees must be assessed against the performance criteria specified in the training package.

Assessment evidence is gathered from a variety of sources. To complete a phase of training trainees are required to:

- undertake individual competency assessments
- complete relevant sections in their learning guide
- undertake additional practical assessments set by the assessor
- satisfy verification requirements where applicable
- where relevant, undertake integrated assessment

Gathering information from a range of sources in this way is seen as a means of ensuring quality of assessment. At least three pieces of evidence are required to confirm competency.

No trainee is assessed in all units by the same assessor, thus reducing the overall likelihood of bias.

Appeals procedure

Regency Institute has an established appeals procedure applicable to assessment. The process used is the standard SA Department of Education, Training and Employment appeals process which, in effect, consists of three stages:

- the student endeavours to resolve the issue through discussion with the assessor
- the student takes the matter to the Retail Studies Co-ordinator or educational manager and may call on Student Services for assistance
- the matter goes to a panel for discussion and resolution

Recognition of current competencies/recognition of prior learning

Recognition of current competency (and recognition of prior learning) is undertaken in accordance with a general process laid down by the Regency Institute School of Business Studies. An RCC information package is available to retail studies trainees wishing to apply. Credit transfer arrangements exist for work completed in secondary schools, other TAFE institutes, and with other registered training organisations.

Authentication of work

The institute is increasingly conducting verification testing to authenticate project work undertaken by students and check underpinning knowledge and understanding. This verification can take the form of oral questions or a simple test containing questions requiring short or even just single word answers. It is

particularly important at the higher levels where students prepare procedures and the like. Less verification is needed at level II because it is mainly assessed by observation; however, students are questioned to check the underpinning knowledge.

Equal opportunity legislation

The equal opportunity policy of the SA Department of Education, Training and Employment applies to all training and assessment conducted in retail studies.

Assessors are very careful about the level of language used in assessment and the environment in which the students are assessed. Any problems which arise are dealt with in conjunction with the Institute Equal Opportunities Officer.

Alternative modes of assessment may be offered where practicable. For example, oral instead of written assessment, or the use of audio or video tapes as an assessment medium.

Components of an assessment

A sample of the School of Business Studies assessment instruments for Certificate II in Retail Operations was examined to see how well they covered the essential components of competency specified in the Training Package for Assessment and Workplace Training.

These components are:

- task skills (performing at an acceptable level)
- task management skills (concurrently managing a number of different tasks that make up a job)
- contingency management skills (dealing with the unexpected)
- job/role environment skills (working with others, working as part of a team)
- transfer of skills (using skills learnt in one situation in a new or different context)

With its emphasis on 'on-job' assessment, either in the workplace or in its comprehensive simulated retail workplace, the Business Services Program is well set up to conduct training and assessment in all the components listed above. Indeed, the range of tasks specified in the assessment guides is such that all but 'transfer of skills' are specifically covered.

Competencies are integrated as a design feature of the training package with many of the competencies overlapping (for example, 'Interact with customers', 'Communicate in the retail workplace' and 'Apply point of sale procedures'). Students/trainees demonstrate transferability of the competencies and skills learned by being assessed in different situations, with different assessors.

Appendix C

A review of the literature on the validity of competency-based assessment

Introduction

There is a relatively small amount of Australian research on validity, especially in the area of competency-based assessment. In their review of the literature on assessment of competency-based training in Australia, Toohey et al. (1995) make only indirect reference to validity issues. Furthermore, most sources (NTB 1992; Hager et al. 1994; Cropley 1995) define validity along the lines that it is: *The extent to which the assessment method measures what it is supposed to measure.*

This definition is apparently a contraction of a much earlier attempt to define validity by Lindquist (1942) in which he stated:

> *The validity of a test may be defined as the accuracy with which it measures that which it is intended to measure, or as the degree to which it approaches infallibility in measuring what it purports to measure.* (p.213)

However as Ebel (1965) pointed out, this conception of validity has shortcomings because it raises the question *What should the test measure?* Definitions which raise substantial questions of this nature are most unsatisfactory.

It is therefore disappointing to find the recently published Training Package of the National Assessors and Workplace Trainers Body (NAWTB, 1999) defines validity in a similar manner as that proposed by Linquist, namely:

> *A valid assessment assesses what it claims to assess; evidence collected is relevant to the activity and demonstrates that the performance criteria have been met.* (p.18)

The training package definition shares the shortcomings of Lindquist's earlier version and further compounds this by adding what presumably is an 'example' of how to conduct a valid assessment namely, *the collecting of relevant evidence that performance criteria have been met.* Basically this addition has the training package defining validity as *content validity;* that is, validity concerned with showing how well the content of the assessment samples the situations (performance criteria) about which conclusions are to be drawn.

Elsewhere in the package under 'rules of evidence', when describing 'valid evidence' there is also the statement:

> *Evidence of competence must cover the broad range of knowledge, skills and the application of such knowledge and skills specified in the Assessment and Workplace Training Competency Standards, assessors need to ensure that the evidence collected*

focuses on the appropriate knowledge and skills specified in the Performance Criteria and Evidence Guides. (p.21)

Again, we can see the authors of the package are clearly relating 'validity' with 'content', in particular, the content of competency standards as specified by the performance criteria. This review presents evidence to show that such an approach to validity is outdated and a new approach is called for.

The NAWTB training package does, however, recognise the fundamental importance of validity in assessment. The package includes validity as one of the four technical principles that underpin assessment, namely, validity, reliability, flexibility and fairness.

This review has a focus on validity but, as will be seen, it is possible to have a wider interpretation of validity which subsumes principles of reliability, flexibility and fairness. Validity is the main assessment game—some would say the only game!

A new definition of validity

Before turning to look at the history of validity it is important to point out that what follows is based on the notion that validity is not an intrinsic property of assessment instruments, but rather it refers to the soundness of the interpretations and uses of the results of an assessment. In other words, the review adopts a different definition of validity to that commonly used in Australia. Gillis and Bateman (1999) were among the first in Australia to recognise the new approach. They do not however, embrace the unitary nature of validity which follows from the new definition.

A recently completed study of the use of types of validity evidence in Australia (Thomson, Saunders & Foyster 2001) was an attempt to give considerations of validity a higher profile than it had previously enjoyed. The authors defined validity as *the extent to which the interpretation and use of an assessment outcome can be supported by evidence.*

By way of example, consider a successful result on a typical driving license test that consists of some theory questions about road rules and safety along with a practical test in driving a car.

In validating this result, the evidence that could be looked for to support the decision would include:

- checking that all the essential road rules and safety issues are covered in the theory test
- checking that the practical test includes the driver performing a suitable range of signals and manoeuvres
- checking that those deemed competent demonstrate, over time, they have continued to be competent drivers
- checking that it is possible to generalise the results to other assessment conditions that were not part of the original test such as ensuring driving ability under night time conditions (that was not tested) is no different from that demonstrated in daytime

- checking that the theory and practical components of the test 'work together' and contribute in similar ways to the result

Gathering evidence in this way enables the assessor to develop a sound agreement to support the interpretation of the result and the relevance of its use.

The Thomson, Saunders and Foyster (2001) definition is a more 'VET friendly' version of that in the latest edition of the *Standards for educational and psychological testing* (AERA 1999) which states that validity is:

> *The degree to which accumulated evidence and theory support specific interpretations of test scores entailed by proposed uses of a test.* (p.184)

This definition uses the word 'test' where, in Australia, we would say 'assessment'. 'Tests' and 'testings' have never recovered from the bad reputation given to them by educators in this country more than 20 years ago. The reasons for this are not relevant to the review and it is enough to note that, for the most part, 'test' and 'assessment' mean the same thing and will be used interchangeably in this context. The Americans see assessment as a broader term which integrates test information with information from other sources, (AERA 1999, p.3). However, when considering the application of validity evidence, no distinction is made between 'tests' and 'assessments'.

The evolution of validity

The standards for people whose job it has been to develop tests or assessment instruments have a long history. Since 1954 the American Psychological Association (APA), the American Educational Research Association (AERA) and the National Council on Measurement in Education (NCME) have published standards for educational and psychological tests. There have been six publications in all. From 1966, the standards have been joint publications involving input from all three organisations. The latest revision of the standards was published in December 1999.

In the early 1950s the validity components of the assessment standards were generally expressed in terms of:

- how faithfully a test represented the curriculum content
- how accurately it predicted some specific criteria (for example, future success in a job)

By the mid-1950s the notion of construct validity had emerged. Cronbach and Meehl (1955) were the first to point out that, as well as wanting to know about content and predicability, assessors wanted to know what qualities were being measured, that is, what explanations (constructs) accounted for performance on assessment. And it was these authors that first noted 'One does not validate a test, but only a principle for making inferences' (p.297).

Messick (1989) has traced the evolution of the concept of validity from those early years and he identifies a trend towards limiting the number of validity types from initially, three or four to today's focus on a unitary notion of validity—the assessors' version of one size fits all!

Messick draws on the APA, AERA, and NCME publications to note that, by 1966, there were three traditional validity types, one of which had two sub-divisions, namely:

- *Content validity:* evaluated by showing how well the content of the test samples the class of situations or subject matter about which conclusions are to be drawn
- *Criterion-related validity:* evaluated by comparing the test scores with one or more external variables (called criteria) considered to provide a direct measure of the characteristics or behaviour in question. Criterion-related validity's two sub-divisions are:
  - *Predictive validity:* indicates the extent to which an individual's future level on the criterion is predicted from prior test performance
  - *Concurrent validity:* indicates the extent to which the test scores estimate an individual's present standing on the criterion
- *Construct validity:* evaluated by investigating what qualities a test measures, that is, by determining the degree to which certain explanatory concepts or constructs account for performance on the test.

Messick goes on to note that:

> *Almost any kind of information about a test can contribute to an understanding of its construct validity, but the contribution becomes stronger if the degree of fit of the information with the theoretical rationale underlying score interpretation is explicitly evaluated.* (p.17)

Messick's paper was most influential; it ran to over 100 000 words with more than 500 citations and it was the evidence he presented, supported by other authors (for example, Linn et al. 1991; Cronbach 1971) that has led to validity now being recognised as '… the most fundamental consideration in developing and evaluating tests' (AERA 1999, p.9). This elevation in the status of validity comes at the expense of reliability. As Linn et al. (1991), note 'Reliability has too often been overemphasised at the expense of validity; validity itself has been viewed too narrowly' (p.16). Indeed, Moss (1994) has argued that it is possible to have validity without reliability (although her argument is based on special premises) and Jessup, a UK researcher in this area (1989), suggested that attempts to improve reliability of competency-based assessments can actually reduce validity. He writes:

> *Take an example. If I faithfully observe the performance criteria and make a valid assessment of an individual's competence and you do not, a comparison between us would indicate that the assessment is unreliable. The solution to this problem is not to try and obtain agreement between the two assessments, as this would very likely reduce the validity of my assessment. The solution to this problem, and I suggest all similar problems, is to check whether the assessments conform to the requirements in the elements of competence and their performance criteria, i.e. check the validity. In all circumstances assessments should be checked against the external criterion and not with each other. If two assessments are both valid they will naturally be comparable and thus reliable, but this is incidental.* (p.192)

Jessup's point is a nice one, but it deals with only one example and (it is to be hoped) an unusual one at that. If an individual assessor makes a different

assessment of the same evidence presented at different times, then we have a situation with unacceptable reliability and unknown validity.

Other UK-based researchers have also put the argument for greater emphasis on validity. The UK work is especially interesting as much of it is devoted to vocational assessments whereas this is not the case for the majority of the research based in the USA that has been cited so far.

Indeed, the vocationally oriented UK research on validity was occurring in parallel with that in the USA with its bias towards 'high stakes' standardised assessments. (Wolf & Silver 1986; Nuttall 1986, 1987; Wolf 1995). In his 1986 paper Nuttall concluded:

> *With the benefit of hindsight, it seems strange that so much effort should have been put into the development and validation of general paper-and-pencil tests, when everything points to their artificiality, their remoteness from the nature of any normal job and their unelaborative conditions of administration. It seems likely that considerations of utility and reliability have prevailed over considerations of validity. The signs are now that validity is claiming its rightful pre-eminent position, and that is helping to unleash potential.* (p.33)

Tolley and Murphy (1998) have reviewed the range of approaches to validity in a competency-based assessment system. The wide range of the approaches and definitions they identified led them to advocate further research, particularly in relation to validity and transferability of vocational qualifications.

Constructs

Until very recently, a 'construct' in educational and psychological assessment and evaluation was understood to be some trait or characteristic that was not directly observable. The construct had to be inferred from an observation. Therefore 'competence' was a construct inferred from observations of 'performance'. This view of constructs was not without its critics. The literature on construct validation reviewed by Norris (1983) identifies a whole series of disagreements over the nature of constructs. The AERA's (1999) standards overcome the disagreements to some extent by departing from the historical use of the term. The standards adopt a broader definition of a construct as 'being any concept or characteristic that an assessment is designed to measure'. Under this definition competence and performance are both constructs and in a competency-based system such as exists in Australia all assessment evidence can be related to a concept or characteristic and hence a construct.

This notion of construct will be used in the discussions of validation evidence that follow. But it needs to be appreciated that this is a development of Messick's (1989) position referred to earlier. Unlike the 'evidence-based' approach used in the 1999 standards, Messick separates evidence related to consequences from construct validity as shown in figure 1.

Figure 1: Messick's framework for the validation of assessments

	Result interpretation	Result use
Evidential basis	Construct validity A	Construct validity & relevance/utility B
Consequential basis	Value implications C	Social consequence D

The top row of the figure presents the construct-based concept of validity and the bottom row allows separately for Messick's view that the consequences of how we use the results of an assessment must also be considered.

How can these facets of validity be applied in a competency-based approach? Consider the example (from child care studies) of a performance assessment in bathing a baby where the baby is, in fact, a plastic manikin. An interpretation of each facet (A,B,C & D) of figure 1 can be given as:

A. Firstly, it could be argued that an assessment that ignores the communication skills such as touching, speaking and listening, that normally occurs between bather and baby does not adequately represent the requirement of the training program. This is an argument about the evidential basis of result interpretation (such an assessment would be said to under-represent the construct of 'communication skills').

B. It could also be argued that leaving out the elements of touching, speaking and listening reduces the ability of the assessment to predict a trainee's likely success in dealing with a real baby, which would be an argument about the relevance of the evidential basis of result use.

C. It could certainly be argued that leaving out the above communication skills would send the message that such aspects of baby bathing are not important, thus distorting the values associated with the training program (consequential basis of result interpretation).

D. Finally, it could be argued that unless aspects of communication skills in the bathing process were incorporated into the assessment, then instructors would not teach, or would place less emphasis on, these aspects (consequential basis of result use).

(Adapted from William 1998)

As William (1998) points out, an aspect of a training program that is assessed comes to be seen as more important than one not assessed, resulting in implications for the values associated with the program. For example, if performance of a particular element of competency is not assessed, an assumption may be made that it is less important than other elements. So a consequence of the use of such limited assessments is that instructors place less emphasis on (or ignore completely), those aspects of the training program that are not assessed.

Social and political implications of consequential validity

Other authors put a greater emphasis on the social and political implications of consequential validity than that given in the above example. Garcia and Pearson (1994) highlight the impact that assessments can have on the lives of students and build a case to show how assessments have been used to impede the progress of African–American, Latino and Asian–American students in the United States. For example, they note that:

> *Asian-American students have earned very high quantitative scores on the SAT, especially the nonverbal section, leading to their over representation in admission pools for elite public universities. Suddenly new admissions criteria (e.g. writing samples) have been added, some would argue, in order to prevent their over representation.* (p.357)

Nor do they see that the current approaches towards ensuring assessments are more equitable will necessarily result in improved validity. They point out that the 'level playing field' approach to equity which has students performing the same tasks under the same conditions only perpetuates entrenched differences between 'advantaged' and 'disadvantaged' groups. However, they cite research with a 'best foot forward' approach where 'disadvantaged' students perform very differently when given choice of assessments that includes choice of projects and socio-linguistic contexts in which to work.

It is not surprising that the social and political issues raised by these considerations of consequential validity have generated more than a little controversy.

Consequential validity was the subject of a protracted internet debate over a period of three months in 1996 on the American Educational Research Association's Measurement and Research website. The issue became quite heated with test developers who claimed that any unintended consequences of tests were not their responsibility being compared with weapons manufacturers who made similar claims about misuses of their products. The main arguments revolved around:

- whether or not the consequences of test use should be considered to be an aspect of validity
- just whom should take responsibility for the misinterpretation of assessment results

The new edition of the Standards for educational and psychological testing (AERA 1999) has addressed both these issues along with some of the more socio-political ones referred to by Garcia and Pearson (1994).

The authors of the 1999 standards contend that, although information about the consequences of assessing may influence decisions about assessment use '… such consequences do not in and of themselves detract from the validity of intended test interpretations' (p.16). Any judgement about the validity would merit further investigation.

> *Take, as an example, a finding of different hiring rates for members of different groups as a consequence of using an employment test. If the difference is due solely to an unequal distribution of the skills the test purports to measure, and if those skills are, in fact, important contributors to job performance, then the finding of group*

> *differences per se does not imply any lack of validity for the intended inference. If, however, the test measured skill differences unrelated to job performance (e.g. a sophisticated reading test for a job that required only minimal functional literacy), or if the differences were due to the test's sensitivity to some examinee characteristic not intended to be part of the test construct, then validity would be called into question, even if test scores correlated positively with some measure of job performance.* (p.16)

Therefore, if the evidence about consequences can be directly traced to some property of the assessment instrument, then consequential validity is an issue and one that can be dealt with by the assessors. On the other hand, if evidence about consequences cannot be traced to a property of the assessment instrument, but rather reflects real differences in performance of the groups concerned, then the case is outside the purview of validity.

On this basis the 'level playing field' approach to assessment which Garcia and Pearson (1994) see as 'perpetuating entrenched differences' between certain groups, although crucial to informing policy-makers, would fall outside the purview of validity (assuming, of course, the assessment in question was one of good quality).

Despite the current interest in consequences, it seems that consequential evidence has not had a great impact on current assessment practices. Little has changed since Linn et al. (1991) noted:

> *Although theoreticians such as Messick (1989; see also Cronbach, 1988) have stressed the criticality of giving attention to the consequential basis of validity, prior to the recent pleas for authentic assessment, consequences could rarely be listed among the major criteria by which the technical adequacy of an assessment was evaluated. If performance-based assessments are going to have a chance of realizing the potential that the major proponents in the movement hope for, it will be essential that the consequential basis of validity be given much greater prominence among the criteria that are used for judging assessments.* (p.17)

Performance-based assessment

It needs to be pointed out that the performance-based assessments referred to in the American literature (see Linn et al. 1991; AERA 1999) differ from the way the term is used in Australia. As well as practical work and portfolios, the Americans classify assessments involving constructed responses (for example, short written answers, essays) as performance assessments. This can be traced, in part, to their long-standing obsession with multiple choice testing; to write something more than the letters A, B, C or D when answering a question is therefore seen as a performance. The contrast with the VET sector in Australia where performance usually involves doing any of a wide range of activities (that may include writing the occasional essay) must therefore be borne in mind. It is interesting to note America has been moving away from multiple choice assessment in recent times. Barton and Coley (1994) in their review of testing in American schools concluded:

> *The nation is entering an era of change in testing and assessment. Efforts at both the national and state levels are now directed at greater use of performance assessment, constructed response questions, and portfolios based on actual student work.* (p.3)

Who will be responsible for validity?

There is little clarity about just who should take responsibility for delivering the various elements of validity evidence. In the United States there are many candidates—the test developers, the publishers and marketers of tests, those who deliver and administer the tests, the decision makers who use the test results, boards and government agencies as well as state and federal governments.

The Australian candidates for taking responsibility for the validity of competency-based assessment are not so numerous. Here the bulk of the responsibility for getting validity right will inevitably fall on the people who develop and administer the assessment instruments and judge performance. In most competency-based environments such as training colleges and companies this means a small number of individuals, and sometimes a single person. What is more, taking responsibility has legal ramifications for those involved (for example, the duty of care requirements that come with OH&S assessment)

The challenges this presents in terms of revisions to the current assessor training programs and associated resource development are formidable.

Ensuring validity

It is clear that the concept of validity has evolved since Messick's seminal work of 1989 and that the process of evolution continues. Practitioners and theoreticians in assessment are no longer referring to different types of validity but instead to different lines of validity evidence. The lines of validity evidence identified in the 1999 Standards are:

- evidence based on test content
- evidence based on response processes
- evidence based on internal structure
- evidence based on other variables (such as predictability and generalisability)
- evidence based on consequences of testing

The Australian research of Thomson, Saunders and Foyster (2001) referred to earlier had begun before the publication of the 1999 standards. The project addressed eight types (facets) of validity evidence that have been identified by Nitko (1996). These types are based on Messick's (1989) work and that of Linn, Baker and Dunbar (1991) and vary only a little from the 'lines of validity evidence' identified in the 1999 Standards.

Nitko's eight types of validity evidence:

1. Content representativeness and relevance (called content evidence)
2. Types of thinking skills and processes required (called substantive evidence)
3. Relationships among the assessment tasks or parts of the assessment (called internal structure evidence)
4. Relationships of assessment results to the results involving assessment of other variables (called external structure evidence)
5. Reliability over time, assessors and content domain (called reliability evidence)
6. Generalisability over different types of people, under different conditions, or with special instruction/intervention (called generalisability evidence)
7. Value of the intended and/or unintended consequences (called consequential evidence)
8. Cost, efficiency, practicality, instruction features (called practicality evidence)

(Nitko 1996)

The Australian researchers found that Nitko's facets of validity had varying levels of relevance in a competency-based assessment system. Participants in the case studies had given little attention to *consequential evidence* and, when considering *generalisability evidence*, there was a tendency to believe 'all was well', but little attempt to gather the supporting evidence.

Overall, however, there was a general awareness of the importance of the other facets of validity. But in order to give validity the status it deserves, a need for changes to be made to the assessment components of the training packages was identified. One outcome of the project was a diagnostic tool for use in pre-service and in-service training of assessors which provided guidance on how best to deal with validity evidence.

What lessons are there for Australia's VET system

It is interesting to reflect upon what impact adopting the new approach to validity might have on competency-based training and assessment in Australia.

The improved quality of assessment which is an outcome of greater attention being paid to validity evidence must inevitably affect technical aspects of the assessment system. Foremost among these is likely to be a demand for better trained and better prepared assessors.

Assessors will need to give greater attention to the quality of their assessment instruments and administration procedures as well as working to ensure they assess in a reliable and fair way.

To ensure this happens, the existing audit procedures of the training authorities in the States and Territories will need to be enhanced to enable improved scrutiny of the assessment processes that cover the facets of validity evidence.

Much remains to be done in this area as evidenced by submissions to the recent Senate committee enquiry into the quality of vocational education and training in Australia (Senate Committee 2000). The enquiry identified concerns and criticisms that related to '... the lack of consistency, rigour and integrity of processes for registration, performance monitoring and auditing of providers' (p.122).

Actions to correct these problems will need to go hand in hand with the adoption of a new approach to validity.

References

AERA (American Education research Association) 1999, *Standards for educational and psychological testing*, AERA, Washington DC.

ANTA (Australian national Training Authority) 1999, *National Assessment Principles*, Melbourne, ANTA (March).

APA (American Psychological Association) 1966, *Standards for education and psychological tests and manuals*, APA, Washington DC.

Barton, PE & Coley, RJ 1994, *Testing in America's schools*, Educational Testing Service, Princeton NJ.

Cronbach, LJ 1971, 'Test validation' in *Educational measurement*, 2nd ed, ed RL Thorndike, ACE, Washington DC.

Cronbach, LJ & Meehl, PE 1955, 'Construct validity in psychological tests', *Psychological Bulletin*, pp.55, 281–302.

Cropley, M 1955, 'Validity' in *Key aspects of competency-based assessment*, ed WC Hall, Adelaide, NCVER.

Ebel, RL 1965, *Measuring educational achievement*, Prentice Hall, Englewood Cliffs NJ.

Garcia, GE & Pearson, PD 1994, 'Assessment and diversity' in *Review of research in education*, ed L Darling-Hammond, AERA, Washington DC.

Gillis, S & Bateman, A 1999, *Assessing in VET: Issues of validity and reliability*, Review of research, NCVER, Adelaide.

Hager, P, Athanasou, J & Gonczi, A 1994, *Assessment: Technical manual*, AGPS, Canberra.

Jessup, G 1991, *NCQs and the emerging model of education and training*, Falmer, Brighton UK.

Lindquist, EF 1942, *A first course in statistics*, Houghton Mifflin, Boston.

Linn, RL, Baker, EL & Dunbar, SB 1991,' Complex performance-based assessment: Expectations and validation criteria', *Educational Researcher*, vol.20, no.8, pp.5–21.

—— 1994, 'Performance assessment: Policy promises and technical measurement standards', *Educational Researcher*, vol. 23, no.9, pp.4–14.

Messick, S 1989, 'Validity' in *Educational Measurement*, 3rd ed, ed RL Linn, Onyx Press, Phoenix AZ.

Moss, PA 1994,'Can there be validity without reliability?', *Educational Researcher*, March, pp.5–12.

NAWTB (National Assessors' and Workplace Trainers' Body) 1999, *Training package for assessment and workplace training* (BSZ98), ANTA, Melbourne.

Nitko, AL 1996, *Educational assessment of standards*, 2nd ed, Prentice Hall/Merrill Education, Englewood Cliffs NJ, pp. 44–5.

Norris, SP 1983, 'The inconsistencies at the foundation of construct validation theory', in *New directions for program evaluation*, ed ER House, Jossey-Bass, San Francisco.

Nuttall, D 1986, 'Assessments and their characteristics', appendix to *Work based learning*, by A Wolf & R Silver, R&D report 33, Manpower Services Commission.

Nuttall, D 1987, 'The validity of assessment', *European Journal of Psychology of Education*, vol.2, pp.109–18

NTB (National Training Board) 1992, *National competency standards policy and guidelines*, 2nd ed, NTB, Canberra.

Senate Committee (Senate Employment, Workplace Relations, Small Business and Education References Committee) 2000, *Aspiring to excellence: Report into the quality of vocational education and training in Australia: Overview and recommendations*, Canberra.

Tolley, H & Murphy, R 1998, *The validity and transferability of S/NVQs in the Workplace: Report for the DfEE*, School of Education, University of Nottingham, Nottingham.

Toohey, S, Ryan, G, McLean, J & Hughes, C 1995, Assessing competency-based education and training: A literature review, in *Australian and New Zealand Journal of Vocational Education Research*, vol.3, no.2, pp.86–117.

Thomson, P, Saunders, J & Foyster, J 2001, *Improving the validity of competency based assessment* NCVER, Adelaide.

William, D 1998, *Construct-referenced assessment of authentic tasks: Alternatives to norms and criteria*, paper presented at 24th annual conference of the International Association for Educational Assessment, Barbados, May 1999.

Wolf, A & Silver, R 1986, *Work based learning: Trainee assessment by supervisors Manpower Services*, Commission R&D Report 33.

Wolf, A 1995, *Competency-based assessment*, Open University Press, Buckingham.

Appendix D

Procedure for interviews

1. Preliminary contact

a. Briefly describe project

- ask who would be the most appropriate person to approach (if not the person being spoken to)
- ask for a brief outline of the units or modules the organisation uses
- ask what modes of training and assessment are used (on-job, off-job, assessment of current competencies/recognition of prior learning)
- ask what standards the assessors work to (industry standards or their own).

b. Briefly outline information which will be needed for review and analysis (much of this to be viewed or obtained during first interview)

Training programs

- details of training and assessment programs/packages
- details of standards or other requirements on which they are based
- indication of which modules or units it would be best to study

Assessment instruments

- modes of assessment (outside workplace, in workplace on-job, in workplace off-job, recognition of current competency, external assessor, internal/ company assessor)
- examples of assessment instruments
- model answers
- criteria used to determine competency
- number of repeat assessments permitted

Formal recognition of assessments

- information regarding formal recognition or acceptance by training sector (training provider, relevant authorities, etc)
- information regarding formal recognition or acceptance by industry (employers, employer bodies, ITABs, etc)

Details of assessment results

- results over period of 3 years (if possible)
- results for at least 30 learners for each assessment task (if possible)

Information for each assessment reviewed

- number of students assessed
- number deemed competent on first assessment
- number not competent on first assessment
- number repeating assessment and deemed competent
- number not competent on subsequent assessment
- details of individual learner performance on each assessment task
- length of assessment (hours)

Estimated costs of assessment

- cost per student assessment (including cost of assessor's time, cost of assessee's lost work time and materials)
- cost per student for relevant training to prepare students for assessment

c. Arrange interviews

- two interviews preferred if schedules permit

d. Forward information

- project summary
- assessor profile questions
- interview questions for assessors
- interview questions for company as a registered training organisation
- assessment review questions

2. Second contact—*interview #1 (face-to-face interview if possible)

Obtain or view

- training and assessment programs/packages
- standards or other requirements on which they are based
- assessment instruments
- model answers

Obtain

- details of assessment results (as above)
- information for each assessment (as above)

Provide

- copies of questions to be answered

3. Assessment review process

- examine materials and information received or notes on material viewed
- prepare additional questions to clarify or extend information received (the above work is done by a team member using the Assessment Review Process outlined below)
- provide preliminary feedback

4. Third contact—*interview #2 (face-to-face interview)

- obtain answers to questions previously supplied (assessor profile questions, interview questions for assessors, interview questions for company as a registered training organisation)
- provide preliminary feedback

5. Develop case study report

- analyse information and write case study report

6. Follow-up contact

- provide report on findings relevant to the participating organisation and/or assessor and negotiate for information to be approved for publication in project report

Note: * These two interviews can be combined (with suitable modification of affected procedures) if circumstances do not allow two separate interviews.

Assessment review process

Introduction

The purpose of the review is to determine how faithfully the assessment processes reflect the industry standards as laid down in the training packages (or company program/curriculum).

The review process involves members of the research team working through the components of the standards and cross-checking these against the assessment 'instruments' being used by the assessors to judge competency.

As well as this matching exercise, the researchers will make judgements about the components of the assessments and the assessment instruments themselves. In the case of the assessment components, a judgement is made against the published assessment guidelines. (Note, some of these guidelines are already covered in the interview schedule being used with the assessors and these will not be repeated here.) In the case of the assessment instruments, another group of judgements will be made by applying a set of criteria to the instruments.

Matching industry standards components against assessments

Component	Questions	Answers & comments
Performance criteria of the relevant elements of competency	• are all the performance criteria covered?	
Range of variables	• do the assessments cover the range of situations and equipment	
Evidence guide	• are all the critical aspects covered? • are underpinning knowledge, skill and attitudes formally addressed? • are 'What if ...' questions used?	
Key competencies	• are key competencies assessed separately or is acquisition of key competencies assumed through 'application' type evidence?	

Requirements of industry assessment guidelines

Items	Answers & comments
1. Has a database of qualifications awarded and statements of attainment issued been developed? If so, how long has it been in use?	
2. Have the assessment procedures of the company/ organisation ever been audited? If so, when and by whom was this last done?	
3. Do the assessments have an appeals process in place as set down in the guidelines?	
4. Are recognition of current competencies/ recognition of prior learning procedures allowed as part of competency assessment?	
5. Are 'integrated competency assessments' (ICAs) used? (Refer Retail Training Package)	
6. If ICAs are used do they adopt the guideline principle of multiple assessments, i.e. minimum of 3 pieces of evidence?	
7. Do the assessments gather enough evidence, i.e. are enough contexts covered and is consistency of performance taken into account?	
8. Are steps taken to authenticate any work samples or other evidence presented for assessment that was not done in the presence of the assessor?	
9. Do the assessments have provisions for dealing with the requirements of 'equal opportunity' legislation?	

Assessment instruments: Components
Do the assessment instruments, taken together, cover ALL of the required components?

Components	Answers & comments
1. Task skills (performing at an acceptable level)	
2. Task management skills (managing a number of different tasks that make up a job)	
3. Contingency management skills (dealing with the unexpected)	
4. Job/role environment skills (working with others, working as part of a team)	
5. Transfer skills (using skills learnt in one situation in a new or different context)	

Interview questions for registered training organisations

Note: The order and numbering of the questions here are different from that in the report.

Questions	Answers & comments
1. Do the assessments used in your company cover all or only some of the content of your training program? (Probe)	
2. Are your company's assessments recognised by all relevant groups and individuals (management, trainers, assessors and trainees) as being appropriate? Which companies or organisations? How do you know this?	
3. Have you conducted any investigations to check how cost-effective your assessment are? (Probe)	
4. Can you say with confidence that your assessments consistently reflect work practice? Do you think any additional assessment is needed? If so, what would this be? Could any assessments or parts of assessments be omitted? Does the emphasis or balance in the assessment match the emphasis 'on the job'? Is the assessment 'up to date'? Are the assessment tasks worthwhile in themselves? (e.g. do they contribute to learning)?	
5. Are the outcomes of your assessment independent of the timing of the assessment? Do you have to pick the right time? Have you ever been forced to conduct an assessment at a time you considered inappropriate?	
6. Where different assessment tasks have some overlap in content, are the assessment outcomes from the different tasks consistent? Is the interpretation of the outcomes of the assessment tasks unambiguous? What do you do if the outcome from a particular task is ambiguous? (e.g. if repeated attempts are allowed, how many repeats do you allow?)	
7. Are the outcomes of the assessment tasks consistent with other evidence such as work assessments or third-party verification (testimonials)? Probe: to learn how this other evidence is collected and/or used.	
8. Do the outcomes of the assessments tasks satisfactorily predict employment success in terms of further learning (either on or off the job)? Probe: for examples	
9. Are assessments graded in any way, e.g. to identify those who are 'very competent' rather than just 'competent'? How is this done?	

BEST COPY AVAILABLE

Interview questions for registered training organisations cont.

Questions	Answers & comments
10. Are the outcomes of the assessment tasks independent of the assessor(s)? (e.g. are you confident that another assessor would come up with the same assessment results as you?) Probe: to find out if this has ever been checked out in practice and note the procedures used.	
11. Would it be useful to have equivalent forms of the assessment tasks available? (e.g. to have an equivalent form of a knowledge test that you could use?) Probe: for reasons behind answer.	
12. Can the outcomes for an individual assessee on an assessment task be affected by incentives (to either assessee or assessor), special forms of motivation, etc. If YES, ask for examples and any further comments.	
13. Can you think of any examples where the results of an assessment have been used for purposes other than those for which they were designed? (e.g. an assessment of OHS actually being used as a test of knowledge of English) Probe: for comments	
14. Following on from the above, can you think of any cases where there are social consequences for the assessee, not related to the work-specific purpose of the assessment? Probe: for comment	
15. Continuing in this area, can you think of any assessment tasks where the outcomes may be influenced by factors not strictly relevant to the job, such as ethnicity, gender, socio-economic status, age? (colourblindness) Probe: for comment	
16. Are you satisfied with the assessment skills of the available assessors, in terms of their ability to implement the assessment procedures?	
17. Do the assessment tasks include the application of thinking skills other than rote learning? Probe: for example	
18. What percentage of your assessment tasks require you to assess on a one-to-one basis?	

NCVER

The National Centre for Vocational Education Research is Australia's primary research and development organisation in the field of vocational education and training.

NCVER undertakes and manages research programs and monitors the performance of Australia's training system.

NCVER provides a range of information aimed at improving the quality of training at all levels.

ISBN 0 87397 714 9 print edition
ISBN 0 87397 715 7 web edition